

[Ausarbeitung eines frei gehaltenen Vortrags am 17.9.1997 auf der Tagung "Rationalität, Realismus, Revisionen, GAP3" der Gesellschaft für analytische Philosophie in München].

## Anders Zweifeln

### Eine metaphysische Provokation

*Olaf L. Müller*

*Humboldt Universität zu Berlin*

ZUSAMMENFASSUNG. Betrachten Sie folgendes philosophische Gedankenexperiment (das den Matrix-Filmen zugrundeliegt): Sie sind ein "Gehirn im Tank", d.h. Sie haben keinen vollständigen menschlichen Körper, sondern nur ein Gehirn, das in einem Tank mit Nährflüssigkeit schwebt; die Nerven-Enden Ihres Gehirns sind mit einem Super-Computer verbunden, der Ihnen die Illusion verschafft, dass alles in Ordnung sei. Putnams externalistischer Beweis gegen diese Hypothese vom Gehirn im Tank ist wasserdicht. Um den Beweis zu stützen, verteidige ich ihn gegen folgenden starken Einwand: Das eingetankte Gehirn könnte den Beweis Wort für Wort wiederholen. Der Fehler dieses Einwandes liegt darin, dass der tanksprachliche Beweis (wenn richtig verstanden) aus der Tanksprache in unsere übersetzt werden muss und dann gar nicht von Gehirnen und Tanks handelt (sondern vom Speicherinhalt des Simulationscomputers). Meiner Ansicht nach sollten wir uns nicht darüber den Kopf zerbrechen, ob wir ein Gehirn im Tank sind, sondern darüber, ob wir in einer analogen Situation stecken, um eine Ebene nach oben verschoben. Das ist ein metaphysisches Problem, kein erkenntnistheoretisches. Doch auf welche Weise sollten wir die Metaphern in meiner Formulierung des Problems auspacken? Wie so oft laufen wir in der Metaphysik Gefahr, Unsinn zu produzieren, wenn wir nicht aufpassen. Um der Gefahr zu entkommen, schlage ich vor, dem Problem folgende präzise Form zu geben: Was müsste ein eingetanktes Gehirn in seiner Sprache sagen, wenn es seine eigene Situation treffend beschreiben wollte? Es braucht dafür semantisch stabile Ausdrücke. (Ein Ausdruck ist semantisch stabil, wenn er bei Übersetzung aus der Tanksprache in unsere Sprache nicht verändert zu werden braucht; solche Ausdrücke sind vom Externalismus à la Putnam nicht betroffen). Zeitliche Ausdrücke sind semantisch stabil; ebenso die Ausdrücke "Entität", "Art", "andersartig". (Im Unterschied hierzu sind die Ausdrücke "*materieller* Gegenstand" und "*natürliche* Art" semantisch instabil). Diese sprachphilosophische Übung führt mich zu folgender Hypothese, gegen die Putnams wasserdichter Beweis nichts ausrichten kann: "Vielleicht gibt es zeitliche Entitäten, die von anderer Art sind als alle materiellen Gegenstände". Sollte diese Hypothese zutreffen, änderte dies nichts an der Zuverlässigkeit unserer alltäglichen und naturwissenschaftlichen Wissensansprüche.

ANMERKUNG. Dies bislang unveröffentlichte Arbeitspapier lag Vorträgen zugrunde, die ich bei verschiedenen Gelegenheiten ab 1997 frei gehalten habe. Das Papier erscheint hier unverändert in der Form aus dem Jahr 1997, allerdings mit neuem Layout.

## **Anders Zweifeln**

### **Eine metaphysische Provokation**

Indem Descartes den radikalen Zweifel in die Welt hineinmeditierte, versetzte er der theoretischen Philosophie des Abendlandes einen Schock, dessen Ausmass er selbst gewaltig unterschätzt hat. Wie besessen haben die Philosophen nach ihm versucht, ihrer Irritation über den radikalen Zweifel herrzuwerden. Beispielsweise durch Gottvertrauen (indem sie aus Gottes Güte eine Wissensgarantie ableiteten); oder durch Narkose (indem sie den Zweifel ruhigstellten, so als sei überhaupt nichts geschehen); oder durch Oropax (indem sie die skeptischen Argumente nicht zu verstehen vorgaben); oder durch Dumping (indem sie unsere althergebrachten Wissensansprüche so weit herunterschraubten, bis sie nicht mehr wiederzuerkennen und dann in der Tat gegen den cartesischen Bazillus immun waren).

Trotz aller philosophischen Raffinesse, die man in solche Manöver gesteckt hat, liess sich die radikale Skepsis nicht aus der Welt schaffen; die Philosophen sind ihr immer nur ausgewichen. Gibt es denn keine direkte Antwort auf die skeptische Herausforderung? Lange Zeit wollte es so scheinen, bis plötzlich im Jahr 1981 das Unerwartete geschah: Hilary Putnam präsentierte dem staunenden Publikum die glatte Widerlegung eines typischen skeptischen Gedankengangs. Putnams Argument zufolge ist es ausgeschlossen, dass unsere Gehirne seit jeher in einem Tank mit Nährflüssigkeit herumschwimmen und an einen gigantischen Supercomputer angeschlossen sind, der uns die Existenz der äusseren Welt (einschliesslich unserer Körper) täuschend echt vorgaukelt.

Wie wir sehen werden, ist Putnams Argument gegen diese skeptische Hypothese verblüffend einfach: so einfach, dass man sich sofort betrogen fühlt. Dieser blitzartige Reflex unserer philosophischen Intuition treibt die meisten von uns in die Suche nach versteckten Fehlern, Zirkularitäten usw. Die Suche hat wenig Aussicht auf Erfolg; meiner Ansicht nach argumentiert Putnam schlüssig, nicht-zirkulär und mit richtigen Prämissen. Ich gehöre also zu der kleinen Minderheit der Philosophen, die sich von Putnam haben überzeugen lassen. Was aber sollen wir mit unserem instinktiven Misstrauen gegen Putnams Argument anfangen? Sollen wir es dadurch zum Verstummen bringen, dass wir das Argument jeden Abend vorm Schlafengehen auswendig aufsagen? Besser nicht. Vor jeder guten Therapie kommt die Diagnose. Wir sollten zu verstehen versuchen, *warum* sich unser philosophischer Instinkt gegen Putnams Argument wehren will. Die Gegenwehr könnte ja einer tiefen Einsicht entspringen.

Dieser Vermutung möchte ich in den folgenden Überlegungen nachgehen. Meine These lautet: Unser instinktiver Widerstand gegen Putnams Argument beruht auf einem berechtigten metaphysischen Frage-Bedürfnis, und zwar auf einem Bedürfnis, das allein durch radikalen Zweifel gestillt werden kann und uns daher sehr leicht in die Arme des cartesischen Zweifels treibt. Solange wir keine radikal zweiflerische Alternative zum cartesischen Zweifel sehen, müssen wir dessen Widerlegung mit instinktivem Misstrauen begegnen – oder wir müssen das metaphysische Frage-Bedürfnis kaltstellen, das dies Misstrauen ausgelöst hat: Ein Dilemma hält uns gefangen. Wir können uns aus seinem Würgegriff befreien, sobald wir sehen, wie sich das metaphysische Frage-Bedürfnis durch einen uncartesischen, aber radikalen, Zweifel Luft verschaffen kann.

Anders zweifeln, lautet also das Motto, in dessen Namen ich unseren metaphysischen Instinkt mit Putnams anti-cartesischem Argument versöhnen will. Der "andere" (uncartesische, aber radikale) Zweifel, für den ich hier plädieren möchte, unterscheidet sich in vielerlei Hinsicht vom überkommenen radikalen Zweifel à la Descartes. Betrachten wir fünf dieser Unterschiede in einer Art programmatischen Vorschau, bevor wir uns in die Details der philosophischen Auseinandersetzung stürzen!

Erstens unterscheidet sich der andere Zweifel vom cartesischen Zweifel hinsichtlich der erkenntnistheoretischen Auswirkungen. Wer cartesianisch zweifelt, verliert jede Hoffnung auf Wissen um die Aussenwelt und hält somit am Ende weniger in den Händen als vor Beginn seiner skeptischen Meditation. Hingegen nimmt uns der andere Zweifel nichts weg; unseren Anspruch auf Wissen um die Aussenwelt lässt er unangetastet. Trotzdem ist der andere Zweifel nicht weniger radikal als der Zweifel à la Descartes: Er lehrt uns, dass unser Wissen auf dramatische Weise *unvollständig* sein könnte. Nicht die *Berechtigung* unserer Wissensansprüche hinsichtlich der Aussenwelt stellt er infrage, sondern deren *Reichweite*. Wollte man den anderen Zweifel auf leicht paradoxe Weise anpreisen, so könnte man sagen: Wer anders zweifelt, hält am Ende mehr in Händen als vor Beginn seiner skeptischen Meditation – denn einerseits darf er sein Aussenweltwissen behalten (er verliert also nichts), und andererseits gewinnt er ein ganzes Bouquet neuer Fragen hinzu, klarer metaphysischer Fragen, die sich nicht rational beantworten lassen, wie die Frage nach unserer Stellung im Weltganzen oder nach der Reichweite unseres Wissens. Kurz, er wird um Fragen bereichert, ohne Wissen einzubüssen.

Der zweite Unterschied zwischen den beiden Formen des radikalen Zweifels betrifft ihr Verhältnis zur Lebenspraxis. Weil der andere Zweifel die Berechtigung unserer Wissensansprüche hinsichtlich der Aussenwelt gar nicht infrage stellt, kann er unseren tagtäglichen und treffsicheren Umgang mit

ebendieser Aussenwelt nicht behindern. Der andere Zweifel kann gelebt werden. Im Gegensatz dazu lässt sich der cartesische Zweifel im Leben nicht durchhalten. Er kann nur in der staubigen Luft des philosophischen Klassenzimmers überleben, als akademisches Abstraktum, abgetrennt von allem Alltag.

Dies führt uns, drittens, zum unterschiedlichen Nutzen, den die beiden Formen des radikalen Zweifels mit sich bringen. Der cartesische Zweifel ist, wegen seiner Unverträglichkeit mit unserem Leben im Alltag, bestenfalls nutzlos, wenn nicht schädlich. Dass wir ihn krampfhaft aus unseren Unternehmungen ausgrenzen müssen, überschattet die positive Rolle, die das radikale Zweifeln in unserem Leben spielen kann. Der andere Zweifel befreit dies positive Potential. Er kann, behaupte ich, unser Leben verbessern. Er sorgt für Gelassenheit, stimmt uns heiter und weitet unseren spirituellen Horizont.

All diese verlockenden Früchte des anderen Zweifels gibt es natürlich nicht umsonst. Damit sind wir beim vierten Unterschied zwischen den zwei Zweifeln: bei den Kosten. In den cartesianischen Zweifel braucht man nicht viel zu investieren; er kommt kostenlos ins Haus. Er drängt sich uns geradezu auf, und zwar – entgegen aller Dementis seitens der *ordinary language philosophy* – mit ganz alltäglichen sprachlichen Mitteln. Der andere Zweifel kostet mehr: Um ihn dingfest zu machen, werden wir die Ausdrucksfähigkeit unserer Sprache bis an die Grenzen ausreizen müssen. Denn die skeptischen Hypothesen, die uns in den anderen Zweifel leiten sollen, werden sich nicht alltagssprachlich fassen lassen; wir brauchen ein handfestes Stück metaphysischer Sprache, das allererst verständlich gemacht werden muss. Das ist die Bürde, die unsere Überlegungen werden schultern müssen.

Der andere Zweifel ist also teurer als sein cartesisches Gegenstück. Dafür hält er auch mehr aus als dieser: Das ist der fünfte Unterschied zwischen den zwei Zweifeln. Während der cartesische Zweifel durch Putnams Argument schnurstracks ausgeschaltet wird, muss der Gegner des anderen Zweifels stärkeres Geschütz auffahren. Er muss Putnams Argument mit zusätzlichen Annahmen anreichern, deren Geltung strittig sein wird. – Damit will ich nicht sagen, dass der andere Zweifel unangreifbar wäre. Ich sage nur, dass er weniger leicht anzugreifen ist als der überkommene cartesische Zweifel.

Wie gelingt es denn nun, den cartesianischen Zweifel aus der Welt zu schaffen? Mit dieser Frage wollen wir den programmatischen Vergleich zwischen den zwei Zweifeln beiseitelegen und in die argumentative Auseinandersetzung eintreten. Ich werde zunächst (in Abschnitt I) Putnams

Argument darstellen und gegen einen naheliegenden Einwand verteidigen. Dann werde ich (in Abschnitt II) unserem instinktiven Misstrauen gegen das Argument Gehör verschaffen und verschiedene Versuche verwerfen, das Misstrauen auf informative Weise zu artikulieren. Im letzten Abschnitt (III) werde ich eine radikal skeptische Hypothese formulieren, die sich mit Putnams Argument vereinbaren lässt, trotzdem unserem instinktiven Misstrauen gegen das Argument gerecht wird und in den anderen Zweifel führt. Da sich diese Hypothese einer genuin metaphysischen Sprache bedient, werde ich den Verdacht ihrer Sinnlosigkeit dadurch zurückweisen, dass ich ihre mögliche Bedeutung für unser Leben herausarbeite.

### I.

Wir werden, Putnam folgend, exemplarisch argumentieren und eine einzige – typisch cartesische – skeptische Hypothese widerlegen: nämlich die Hypothese vom Gehirn im Tank.<sup>1</sup>

Die Unhaltbarkeit dieser Möglichkeit ergibt sich nach einem einzigen Zwischenschritt aus folgenden beiden Prämissen:

- (1) In der Sprache, in der ich hier gerade schreibe, referiert das Wort 'Tiger' auf Tiger.
- (2) In der Sprache eines Gehirns im Tank referiert das Wort 'Tiger' *nicht* auf Tiger.

Die erste Prämisse stammt aus der Logik; sie ergibt sich aus dem Disquotationsprinzip, das allseits anerkannt wird. Für die zweite Prämisse müssen dagegen deutlich stärkere, inhaltliche Überlegungen mobilisiert werden. Sie drückt die plausible Ablehnung dessen aus, was Putnam "magische Theorie der Referenz" nennt. Um auf einen Gegenstand referieren zu können, muss der Sprecher laut Putnam in irgendeiner kausalen Verbindung mit dem Gegenstand (oder zumindest mit Gegenständen derselben Art) stehen; ohne eine solche kausale Verbindung wäre das Bestehen der fraglichen Referenzrelation ein unerklärliches Mysterium. So weit die negative Doktrin, die zur Stützung der Prämisse (2) bereits ausreicht. Mithilfe welcher positiven Theorie der Referenz diese Ablehnung der magischen Referenztheorie vervollständigt wird, spielt für Putnams eigentliches Argument zunächst einmal keine Rolle. Schon aus den Prämissen (1) und (2) folgt nämlich:

---

<sup>1</sup> Vergl. Putnam [RTH]:1-21. Ich folge der Rekonstruktion von Crispin Wright in [PPTW]. Wie sich das Argument etwa auf die Traumhypothese oder die Möglichkeit des täuschenden Dämons übertragen lässt, kann ich hier nicht erörtern.

(3) Die Sprache, in der ich hier gerade schreibe, ist verschieden von der Sprache eines Gehirns im Tank.

Dieser Schluss scheint mir zwingend zu sein. Die beiden Sprachen haben unterschiedliche (semantische) Eigenschaften; also sind sie nicht identisch. Nun ergibt sich aus (3) sofort:

(4) Ich bin kein Gehirn im Tank.

Denn wäre ich ein Gehirn im Tank, so spräche ich auch die Sprache eines Gehirns im Tank – im Widerspruch zu (3).

Das Argument hat natürlich nicht gezeigt, dass es keine Gehirne im Tank geben kann. Selbst wenn unsere Techniker sie noch nicht konstruieren können, bleibt die Entstehung von Gehirnen im Tank irgendwo im Andromeda-Nebel denkbar: Sie könnten beispielsweise durch einen gigantischen kosmischen Zufall entstehen. Das provoziert folgenden Einwand gegen unser Argument: Das Argument könnte wortwörtlich und mit derselben Überzeugungskraft von einem Gehirn im Tank *wiederholt* werden – mit einer dann falschen Konklusion. Muss somit nicht irgendwo in dem Argument ein Fehler stecken?

Nein. Denn wenn das Argument von einem Gehirn im Tank vorgetragen wird, dann  *klingt*  es zwar wie ein Argument in meiner Sprache; in Wirklichkeit argumentiert das eingetankte Gehirn jedoch in einer Sprache, die meiner Sprache nur syntaktisch gleicht, nicht aber semantisch. Die Wörter der Tanksprache referieren auf ganz andere Gegenstände als ihre (syntaktisch identischen) Gegenstücke aus meiner Sprache. Dass das Wort 'Tiger' in der Tanksprache nicht auf Tiger referiert, haben wir uns bei der Motivation der zweiten Prämisse bereits klargemacht. Dem eingetankten Gehirn geht der nötige kausale Kontakt mit Tigern ab; es ist nie mit Tigern in Berührung gekommen: Wie sollte es da, ohne Zauberei, von Tigern sprechen können? Genausowenig kann es natürlich von Bären, Eulen oder Känguruhs sprechen; in der Tanksprache referieren die Wörter 'Bär', 'Eule', 'Känguruh' daher nicht auf Bären, Eulen bzw. Känguruhs. Und per Analogie folgern wir:

(5) In der Sprache eines Gehirns im Tank referiert das Wort 'Gehirn' *nicht* auf Gehirne.

(6) In der Sprache eines Gehirns im Tank referiert das Wort 'Tank' *nicht* auf Tanks.<sup>2</sup>

---

<sup>2</sup> Einwand: Aber das eingetankte Gehirn steht in kausalem Kontakt mit Tanks – es schwimmt in einem Tank mit Nährflüssigkeit herum, ist auf die richtige Tank-Temperatur angewiesen usw. Antwort: Das ist kein kausaler Kontakt

Wenn wir nun Putnams Argument einem eingetankten Gehirn in den Mund legen und auf Schlüssigkeit abklopfen wollen, dann dürfen wir seine Konklusion:

(4) Ich bin kein Gehirn im Tank,

nicht deshalb zurückweisen, weil dieser Satz in unserer Sprache die Lage des argumentierenden Gehirns im Tank falsch beschreibt. Denn aus (5) und (6) ergibt sich, dass das Gehirn mit (4) gar nicht behauptet hat, kein Gehirn im Tank zu sein.

Diese Antwort auf den Einwand hilft aber nur halb. Denn wenn Putnams Argument von einem eingetankten Gehirn wortwörtlich wiederholt werden kann, dann genügt es uns nicht zu erfahren, dass in diesem Fall die Konklusion irgendetwas anderes besagt als ihr Gegenstück in unserer Sprache. Wer Putnams Argument als Beruhigungsmittel in allen erdenklichen Lebenslagen einsetzen möchte, will sicher sein, dass es stets zu einer wahren Konklusion führt, ganz gleich, von wem es vorgebracht wird.

Diesem Verlangen kann Putnam nachkommen. Er darf sich allerdings nicht länger auf rein negative Aussagen über die Referenzobjekte tanksprachlicher Terme zurückziehen (wie in (2), (5) und (6)). Wenn die Konklusion (4) auch in der Tanksprache wahr sein soll, dann muss sich Putnam ein kleines Stück weiter aus dem Fenster hängen und eine Interpretation der tanksprachlichen Wörter vorschlagen: Ohne Interpretation keine Wahrheit.

Worauf referieren die Wörter eines Gehirns im Tank? Worauf referiert das Gehirn im Tank beispielsweise, wenn es 'Tiger' erwähnt? Da seine Tigereindrücke aus einem Computer herkommen, liegt es nahe, die Antwort irgendwo in diesem Computer zu vermuten.<sup>3</sup> Gewisse Konfigurationen von Bits und Bytes in dessen Arbeitsspeicher verschaffen dem eingetankten Gehirn Tigervorstellungen und verleiten es zuverlässig zu der Aussage: 'Da ist ein Tiger'. Wenn diese Aussage schon so schön mit dem Auftauchen ganz bestimmter Bits und Bytes im Computer korreliert ist, dann handelt sie doch wohl von ebendiesen Bits und Bytes: von Bit-Tigern, wie ich der Kürze halber sagen möchte. Sie besagt (aus unserer Sicht), dass sich an der einschlägigen Stelle im Computer ein Bit-Tiger tummelt. D.h. sie ist *wahr*, wenn sich dort ein Bit-Tiger tummelt.<sup>4</sup>

---

der geeigneten Sorte. Das eingetankte Gehirn kann seinen Tank genauso wenig anfassen, sehen oder riechen, wie es Tiger, Eulen, Bären usw. anfassen, sehen oder riechen kann.

<sup>3</sup> So auch Putnam in [CWoB]:287.

<sup>4</sup> Die Rede von Bit-Tigern möchte ich hier propädeutisch verstanden wissen. Wir werden unten einen genaueren Begriff des Bit-Tigers ausbuchstabieren, der sich mit seinem hier verwendeten Vorläufer nur teilweise deckt.



Und schon können wir die erkenntnistheoretischen Sorgen etwaiger Gehirne im Tank beruhigen: Wenn sie sich durch besonders einprägsame und sehr gegenwärtige Tiger-Eindrücke zu der Behauptung:

(7) Da ist ein Tiger,

hinreißen lassen, dann werden sie damit i.A. richtig liegen. Denn ihre Tiger-Eindrücke werden i.A. von genau den Bit-Tigern ausgelöst, deren Präsenz die Behauptung verlangt.

Totale Skepsis hinsichtlich aller Aussagen über die Aussenwelt ist also nicht einmal für eingetankte Gehirne angebracht. Eingetankte Gehirne haben öfter recht, als man ihnen zugetraut hätte. Wie oft? Wenn es stimmt, dass sie i.A. hinsichtlich dessen richtig liegen, was in ihrer Sprache als Anwesenheit von Tigern gilt, dann dürfte sich dieser Punkt auf all ihre zoologischen Beobachtungen übertragen lassen. Sie liegen also auch richtig, wenn sie aufgrund geeigneter Eindrücke zu folgenden Behauptungen kommen:

(8) Da ist ein Bär,

(9) Da ist eine Eule,

(10) Da ist ein Känguruh.

Denn diese Behauptungen handeln nicht von Bären, Eulen oder Känguruhs, sondern von Bit-Bären, Bit-Eulen und Bit-Känguruhs: also von jenen Bits und Bytes im Computer, die dem Gehirn die fraglichen Tiereindrücke verschaffen.

Auch die botanischen Beobachtungen der eingetankten Gehirne werden i.A. stimmen, ja mehr noch: All ihre Beobachtungssätze werden einen beruhigenden Hang zur Wahrheit zeigen. Und damit brauchen die Gehirne im Tank keine Angst zu haben, dass sie sich so massiv täuschen könnten, wie es der cartesische Zweifel gern hätte.

Etwas Überraschendes ist passiert: Wir hatten uns davon überzeugen wollen, dass Putnams Argument selbst dann funktioniert, wenn es von einem eingetankten Gehirn vorgebracht wird. Hierfür mussten wir uns überlegen, wie die Tanksprache in unsere Sprache zu übersetzen ist. Und rein instinktiv haben wir uns in diesem Übersetzungsprojekt zunächst auf Beobachtungssätze



konzentriert. ('Da ist ein Tiger', 'Da ist ein Känguruh' usw.)<sup>5</sup> Durch die gewonnenen Übersetzungen konnten wir den globalen cartesianischen Zweifel *frontal* ausschalten – und zwar überraschenderweise ohne dass wir hierfür Putnams Argument hätten auswerten müssen! Wir können uns sozusagen schon im voraus darauf verlassen, dass die erhoffte anti-cartesische Wirkung der tanksprachlichen Fassung von Putnams Argument gut mit der tatsächlichen erkenntnistheoretischen Lage der eingetankten Gehirne übereinstimmt.

Trotz dieser Vorab-Garantie wollen wir den ursprünglichen Gedanken zuendespinnen und Putnams Argument aus der Tanksprache in unsere Sprache übersetzen. Da das Argument nicht aus Beobachtungssätzen besteht, müssen wir die Übersetzung der Tanksprache weiter ausbauen. Beim Übersetzen der Beobachtungssätze hatten wir nicht sonderlich tief in den Computer hineinsehen müssen: Es kam einzig und allein auf die Konfiguration der Bits und Bytes an, aus denen der jeweils aktuelle sensuelle Input für das eingetankte Gehirn zusammengerechnet wird. (Es kam sozusagen nur auf die Bits und Bytes im "Beobachtungsfeld" an: einem vergleichsweise kleinen Speicherbereich, dessen Inhalt unmittelbar in die rezeptiven Nervenbahnen des Gehirns eingespeist wird).

Wovon spricht das Gehirn aber, wenn es Behauptungen wagt, die über seine unmittelbare Beobachtung hinausgehen, wie z.B. diese hier:

(11) Irgendwo im Hundertmorgenwald lauert ein Tiger.

Dieser Satz handelt nicht von den Bits und Bytes im Beobachtungsfeld, die irgendwelchen Tigereindrücken unmittelbar vorausgehen. (Vielleicht bekommt das eingetankte Gehirn den fraglichen "Tiger" nie zu Gesicht!) Für die Übersetzung solcher Sätze müssen wir Konfigurationen von Bits und Bytes ausserhalb des Beobachtungsfeldes in Betracht ziehen. Und das bedeutet, leider, dass wir uns zuallererst eine genauere Vorstellung der Architektur des Simulations-Computers verschaffen müssen.

Natürlich könnte der Simulations-Computer nach ganz unterschiedlichen Bauplänen konstruiert sein. Ich werde nur eine einzige unter den denkbaren Realisierungen skizzieren. Für die folgende

---

<sup>5</sup> Quines Begriff des Beobachtungssatzes z.B. aus [WO]:40ff. muss nur *geringfügig* an die ungewöhnliche Lage eingetankter Gehirne angepasst werden, bevor er in der üblichen Weise das Anfangsmaterial fürs radikale Übersetzen bereitstellt.

Diskussion kommt es nicht so sehr darauf an, *in welcher Weise* die nötigen kybernetischen Details spezifiziert sind, sondern darauf, *dass* sie hinreichend genau spezifiziert sind.<sup>6</sup>

Je mehr Informationen über unser Universum im Computer codiert sind, desto leichter fällt es, sich vorzustellen, dass er lebensechte Simulationen ausspucken kann. Schöpfen wir also aus dem vollen, und verschaffen wir ihm alles erdenkliche Faktenwissen über einen willkürlich gewählten Zeitpunkt. Genauer: Bis hinunter auf die Ebene der Elementarteilchen sei eine komplette Repräsentation unseres Universums vom 1.4.1947 abgespeichert.<sup>7</sup> Der physikalische Raum wird durch einen gigantischen (und ihm isomorphen) Massenspeicher repräsentiert (den "Universalspeicher"), in dessen Speicherplätzen Codes für die verschiedenen Mitglieder aus dem Teilchenzoo abgelegt sein können. (So werden beispielsweise Elektronen durch Sequenzen der Form '101' codiert; solche Sequenzen – in ihrer konkreten Realisierung im Computer! – werde ich als "Bit-Elektronen" bezeichnen.)

Wir sollten uns von Anbeginn klarmachen, dass physikalisch benachbarte Speicherplätze im Universalpeicher nicht unbedingt physikalisch benachbarte Raumportionen des Universums repräsentieren müssen. Nicht die physikalische Organisation des Universalpeichers in der Hardware sichert ihm die Isomorphie zum physikalischen Raum – es genügt, wenn sich die topologischen Relationen des physikalischen Raumes isomorph auf der Menge der *Namen* der Speicherplätze wiederfinden.<sup>8</sup>

Um die zwei Sorten von Nachbarschaftsverhältnissen nicht durcheinanderzubringen, wollen wir zwei Speicherplätze *kybernetisch benachbart* nennen, wenn sie physikalisch benachbarte Raumportionen im Universum repräsentieren.<sup>9</sup> Die Rede von der kybernetischen Nachbarschaft kann in naheliegender Weise auf die gespeicherten Codes übertragen werden: Zwei Bit-Elektronen, zum Beispiel, sind kybernetisch benachbart, wenn sie in kybernetisch benachbarten Speicherplätzen abgelegt sind.

---

<sup>6</sup> Meiner Ansicht nach mangelt es der Literatur in dieser Hinsicht an der wünschenswerten Genauigkeit.

<sup>7</sup> Für unsere Zwecke sei es erlaubt, vereinfachend anzunehmen, dass jedes Elementarteilchen einen eindeutigen Ort einnimmt und dass die Angabe aller Örter (und aller Geschwindigkeiten) aller Teilchen das Universum genau genug beschreibt.

<sup>8</sup> Die Namen der Speicherplätze könnten z.B. aus den dreistelligen Koordinaten des Mittelpunkts der durch sie repräsentierten Raumportionen gebildet sein. Durch eine bijektive Zuordnung der Raumportionen in die Menge der Speicherplatz-Namen können dann alle interessierenden Relationen zwischen den repräsentierten Raumportionen durch gleichwertige Relationen zwischen Speicherplatz-Namen wiedergegeben werden.

<sup>9</sup> Im Lichte der vorigen Fussnote lässt sich kybernetische Nachbarschaft von Speicherplätzen anhand ihrer Namen feststellen.

Betrachten wir nun ein Bit-Neutron (das ist eine irgendwo im Universalspeicher abgelegte Sequenz der Form '011'), in dessen unmittelbarer kybernetischer Nachbarschaft ein weiteres Bit-Neutron ('011') und zwei Bit-Protonen ('010') abgespeichert sind. Die mereologische Summe der vier Sequenzen wollen wir als "Bit-Heliumkern" bezeichnen: Die Bezeichnung liegt nahe, denn seine Bestandteile repräsentieren die Teile eines echten Heliumkerns, draussen im Universum. In diesem Stil können wir im Universalspeicher den Aufbau der Welt kybernetisch nachbilden. Durch mereologisches Aufsummieren erhalten wir nacheinander Bit-Atome und Bit-Moleküle, Bit-Gase und Bit-Kristalle, Bit-Galaxien und Bit-Planeten, Bit-Tiger und Bit-Känguruhs. Alles, was am 1.4.1947 durchs Universum kreucht und fleucht, findet sich im Universalspeicher als mereologische Summe gewisser (kybernetisch mehr oder minder benachbarter) Codes für Elementarteilchen wieder.

Bislang haben wir den Inhalt des Universalspeichers als stehendes Bild des Universums vom 1.1.1947 aufgebaut, dem Geburtsdatum von Prinz Charles. Das war zufällig der Zeitpunkt einer gigantischen kosmischen Katastrophe, in dessen Verlauf fast unser gesamtes Universum in einem schwarzen Loch verschwand: Übrig blieben ein Tank voll Nährflüssigkeit, darin Prinz Charles' Gehirn, angeschlossen an einen Supercomputer mit einem Universalspeicher, der durch einen aberwitzigen Zufall genau so konfiguriert war, wie wir es uns gerade lang und breit ausgemalt haben. Und man möchte es kaum glauben: Wie es der Zufall wollte, lief in dem Computer ein Programm, das seither den Universalspeicher in Echtzeit laufend aktualisierte, so als repräsentierte er den normalen Fortgang der Geschichte unseres Universums, wie sie ohne die kosmische Katastrophe weitergegangen wäre.<sup>10</sup> Besser noch: Ein gleichfalls zufällig entstandenens Kommunikationsmodul sorgt für den Datenaustausch zwischen Computer und Charles' Hirn: Wann immer das Gehirn elektrische Signale nach draussen sendet, wird an der geeigneten Stelle in der kybernetischen Nachbarschaft des Bit-Hirns ein Bit-Elektron abgelegt: Dadurch exekutiert der Computer die prinzlichen Entscheidungen im Universalspeicher. Umgekehrt werden Bit-Elektronen, die den sensuellen Input des prinzlichen Gehirns repräsentieren würden, hätte Charles seinen Restkörper in der ihm angemessenen Umgebung des Buckingham-Palasts behalten, in Form echter Elektronen auf geeigneten Nervenbahnen ins eingetankte Prinzenhirn hineingeleitet. Dadurch erlebt das Gehirn, was in der kybernetischen Nachbarschaft seines Bit-Körpers vor sich geht. Und diese Erlebnisse gleichen minutiös jenen Erlebnissen, die der heranwachsende Prinz gehabt hätte,

---

<sup>10</sup> Wir nehmen also – wieder zur Vereinfachung – an, dass das Universum deterministisch funktioniert und wirklich Schritt für Schritt weitergerechnet werden kann. Hierzu braucht der Computer neben den Orten aller Teilchen all ihre Geschwindigkeiten und sämtliche physikalischen Gesetze (die irgendwo ausserhalb des Universalspeichers abgelegt sein werden).

wären nicht sein Körper, der Buckingham-Palast und der ganze Rest des Universums am 1.4.1947 für immer in einem schwarzen Loch verschwunden.

Zwanzig Jahre später: Das eingetankte Prinzenhirn hat eine vorzügliche britische Erziehung simuliert bekommen und parliert fließend in allerlei Tanksprachen, u.a. auf Tankdeutsch. Was bedeuten die geflügelten Worte, die seinem Bit-Mund entfleuchen? Die vorhin bereits angedeuteten Übersetzungen tankdeutscher Beobachtungssätze können wir nun genauer fassen. Beispielsweise besagt der tankdeutsche Satz:

(12) Da ist ein Tiger ganz in meiner Nähe,

in unserer Sprache soviel wie:

(13) In der kybernetischen Nähe meines Bit-Körpers ist ein Bit-Tiger abgespeichert.<sup>11</sup>

Und nehmen wir an, der Bit-Prinz beobachtet das Raubtier in Aktion und sagt:

(14) Es naht ein Tiger.

Dann behauptet er damit natürlich nicht, dass sich ein Bit-Tiger immer weiter an seinen Bit-Körper annähert, dass sich also der Abstand zwischen den Speicherzellen, in denen sein Bit-Körper abgelegt ist, und den vom Bit-Tiger eingenommenen Speicherzellen *tatsächlich* verringert. Denn wie wir uns vorhin klargemacht haben, kommt es nicht auf den tatsächlichen (physikalischen) Abstand irgendwelcher Sequenzen im Universalspeicher an, sondern auf ihren *kybernetischen* Abstand. Der Satz besagt also vielmehr folgendes:

(15) Der kybernetische Abstand zwischen meinem Bit-Körper und einem (kybernetisch ohnehin schon bedrohlich nah abgespeicherten) Bit-Tiger sinkt aufgrund kybernetischer Bewegungen des Bit-Tigers durch den Universalspeicher.<sup>12</sup>

---

<sup>11</sup> Ganz in Übereinstimmung hiermit steht Putnams Forderung, dass auch die räumlichen Vokabeln eingetanker Gehirne uminterpretiert werden müssen ([CWoB]:286).

<sup>12</sup> Die letzten acht Worte dieses Satzes sind nötig, um anzuzeigen, dass im fraglichen Geschehen der Bit-Tiger die Initiative ergriffen hat und dass die Abstandsverringering nicht etwa durch bit-prinzliche Bewegungen zustandekommt.

Eine langatmige Formulierung, die wir bequemerweise so abkürzen wollen:

(16) Es naht kybernetisch ein Bit-Tiger.

Ich möchte den Zusatz "kybernetisch" im folgenden immer für derartige Abkürzungen verwenden, ohne jedesmal die langatmige Formulierung ausbuchstabieren zu müssen, die eigentlich angebracht wäre. (Der Zusatz "kybernetisch" zeigt also zuallererst einen nicht-standardgemässen Gebrauch räumlicher Redewendungen an wie in (13), überträgt sich dann auf die Rede von räumlichen *Veränderungen* wie in (15) und (16) und schliesslich ganz allgemein auf die Rede von *Handlungen* irgendwelcher Bit-Wesen, z.B. auf die ganz naheliegende Rede von kybernetischen Küssen).

Wir haben genug Material beisammen, um die Übersetzung *beobachtungsferner* Sätze aus der Tanksprache anzupacken. Erinnern wir uns an unser Beispiel von vorhin:

(11) Irgendwo im Hundertmorgenwald lauert ein Tiger.

Es liegt nahe, es mit folgender Übersetzung zu versuchen:

(17) Irgendwo im kybernetischen Innern des Bit-Hundertmorgenwalds lauert kybernetisch ein Bit-Tiger.

Doch hier erhebt sich ein Hindernis. Unser eingetanktes Gehirn ist vielleicht noch nie in kausalen Kontakt mit genau dem Bit-Tiger geraten, der da im Bit-Gebüsch kybernetisch lauert – kann denn dann das Gehirn überhaupt von diesem Bit-Tiger reden, auf ihn referieren? Haben wir vorhin nicht einen kausalen Kontakt zwischen Sprecher und Referenzobjekt verlangt, als wir die magische Theorie der Referenz im Verein mit Putnam zurückgewiesen haben?

Nein. So streng dürfen wir die referentiellen Fähigkeiten des eingetankten Gehirns nicht beschränken. Wir gehen ja auch mit uns selbst nicht so streng ins Gericht. Wenn wir von irgendeinem lauernden Tiger reden, dann kommen für die Wahrheit unserer Behauptung nicht allein die Tiger in Betracht, denen zu begegnen wir schon die Ehre hatten. In Betracht kommt *jeder* Tiger, bekannt und unbekannt. Wieso ausgerechnet jeder *Tiger*? Die Frage haben Putnam und Kripke in

ihren Überlegungen zur Referenz von Termen für natürliche Arten beantwortet.<sup>13</sup> Ihr Bild<sup>14</sup> sieht so aus: Ein Sprecher – oder eine Gruppe von Sprechern<sup>15</sup> – kann durch kausalen Kontakt mit einer paradigmatischen Auswahl von z.B. Tigern einen Artnamen prägen, der auf alle Objekte referiert, die dieselbe *innere Struktur* haben wie die meisten Exemplare der ursprünglichen paradigmatischen Auswahl. Worauf der fragliche Artnamen referiert, hängt diesem Bild zufolge nicht davon ab, ob der Sprecher die strukturellen Eigenschaften kennt, die den meisten Exemplaren der paradigmatischen Auswahl gemeinsam zukommen. Selbst wenn der Sprecher die ursprüngliche Auswahl anhand irgendwelcher oberflächlicher Gemeinsamkeiten zusammengestellt hat (oder gar aufgrund einer falschen Theorie über angebliche Gemeinsamkeiten der Tiefenstruktur), referiert der Artnamen nicht etwa auf alle Objekte mit denselben Oberflächenmerkmalen wie die ursprünglichen Exemplare (bzw. mit der ihnen fälschlich zugeschriebenen Tiefenstruktur). Vielmehr referiert er auf alle Objekte mit genau der Tiefenstruktur, die den meisten Exemplaren der ursprünglichen Auswahl *wirklich* zukommt – einerlei, ob der Sprecher davon nun weiss oder nicht. Die Umwelt des Sprechers entscheidet sozusagen mit über die Referenz seiner Artnamen.

Was bedeutet das für unser eingetanktes Gehirn? Der Bit-Prinz ist schon in jungen Jahren regelmässig auf die kybernetische Safari mitgekommen. So erfreute sich das eingetankte Gehirn reichhaltiger Tiger-Erfahrungen, die aus seinem engen kausalen Kontakt mit ganz bestimmten Speicherinhalten im Universalspeicher herrührten. Hierdurch wurde die Referenz seines Artnamens 'Tiger' fixiert. Der Artnamen referiert auf alle Objekte, die dieselbe innere Struktur haben wie die Objekte, die vom eingetankten Gehirn ursprünglich tigerartig wahrgenommen worden sind. Worin besteht deren innere Struktur? In einer ganz bestimmten kybernetischen Zusammensetzung aus bit-biologischen Bit-Zellen, in deren kybernetischem Innern Bit-DNS-Moleküle vorkommen, also letztlich eine grosse Zahl abgespeicherter Sequenzen der Form '101', '011', '010', die in hochkomplizierten kybernetischen Nachbarschaftsverhältnissen arrangiert sind. Objekte mit einer solchen Tiefenstruktur dürfen mit Fug und Recht als Bit-Tiger bezeichnet werden, und zwar ganz unabhängig davon, wo sie im Universalspeicher abgelegt sind. Wenn also ein solcher Bit-Tiger im Bit-Gebüsch kybernetisch lauert und noch nie in kausalen Kontakt mit dem eingetankten Gehirn

<sup>13</sup> Vergl Kripke [NN], Putnam [MoM] sowie [IWNH]:60/61.

<sup>14</sup> Beide Philosophen legen Wert auf die wittgensteinianistische Feststellung, dass sie keine *Theorie* aufzustellen wünschen.

<sup>15</sup> Ich werde alle Komplikationen aufgrund der linguistischen Arbeitsteilung ignorieren. Das folgende Bild ist daher nur ein Schwarzweiss-Bild. Es berücksichtigt nur einen einzigen Sprecher – anstelle einer ganzen bunten Sprachgemeinschaft von Experten, Nicht-Experten, Kindern und Clowns.

geraten ist, dann kann das Gehirn trotzdem auf diesen Bit-Tiger referieren: und zwar sogar dann, wenn das eingetankte Gehirn die beschriebene innere Struktur von Bit-Tigern nicht einmal erahnt.<sup>16</sup>

Jedenfalls können wir jetzt endlich die erste positive Aussage zur Referenz eines tanksprachlichen Wortes festhalten:

(18) In der Sprache eines Gehirns im Tank referiert das Wort 'Tiger' auf Bit-Tiger.

Die anderen tanksprachlichen Artnamen funktionieren natürlich genauso:

(19) In der Sprache eines Gehirns im Tank referiert das Wort 'Bär' auf Bit-Bären.

(20) In der Sprache eines Gehirns im Tank referiert das Wort 'Känguruh' auf Bit-Känguruhs.

Auch tanksprachliche Bezeichnungen für Artefakte müssen sich diesem Muster fügen:

(21) In der Sprache eines Gehirns im Tank referiert das Wort 'Tank' auf Bit-Tanks.

(22) In der Sprache eines Gehirns im Tank referiert das Wort 'Computer' auf Bit-Computer.<sup>17</sup>

Wir haben die Übersetzung der Tanksprache weit genug vorangetrieben, um zu überlegen, wieviel die Konklusion aus Putnams Argument wert ist, wenn sie von einem eingetankten Gehirn erschlossen wird. Was meint das Gehirn, wenn es

(4) Ich bin kein Gehirn im Tank,

---

<sup>16</sup> Kann das eingetankte Gehirn diese innere Struktur denn richtig erfassen? Ja. Natürlich wird (und darf) es sie nicht in den Worten beschreiben, die *wir* oben verwendet haben: Es wird (und darf) die Struktur von Bit-Tigern nicht als kybernetische Anordnung irgendwelcher Bits in irgendeinem Universalspeicher beschreiben. Das eingetankte Gehirn hat es einfacher. Es braucht nur zu sagen: 'Ein Tiger [sic] ist ein Tier mit den und den DNS-Molekülen'. Aber selbst diese Beschreibung muss das eingetankte Gehirn nicht zur Verfügung haben, um mit seinem Wort 'Tiger' auf Bit-Tiger zu referieren. Zum Vergleich: Schon vor Entwicklung der Molekular-Biologie konnten Menschen mit ihrem Wort 'Tiger' genau auf Tiger referieren. Oberflächlich ununterscheidbare Objekte mit anderer Tiefenstruktur (z.B. vom Mars ferngesteuerte Tigermaschinen) hätten schon vor zweihundert Jahren keine Behauptung über Tiger wahrgemacht.

<sup>17</sup> Sind *zufällig* entstandene Tanks oder Computer Artefakte? Eher nicht. Aber Referenzregeln nach Art von (21) und (22) gelten sogar für zufällig entstandene Dinge, die nur zufällig wie Artefakte funktionieren.



sagt? Um uns nicht verwirren zu lassen, sollten wir vielleicht das Personalpronomen in ein Possessivpronomen verwandeln und die Konklusion – immer noch in der Tanksprache – folgendermassen fassen:

(23) Mein Gehirn steckt nicht im Tank (und ist nicht an einen Simulationscomputer angeschlossen).

Die deutsche Übersetzung dieses Satzes liegt auf der Hand:

(24) Mein Bit-Gehirn ist nicht im kybernetischen Innern eines Bit-Tanks abgespeichert (und ist nicht kybernetisch an einen Bit-Computer angeschlossen).

Und dieser Satz beschreibt die Lage des Gehirns im Tank richtig. Zwar steckt das *Gehirn* des Gehirns im Tank *in* einem *Tank*; aber sein *Bit-Gehirn* ist im *kybernetischen Innern* seines *Bit-Schädels* abgespeichert – steckt also keineswegs im kybernetischen Innern irgendeines Bit-Tanks und ist auch nicht kybernetisch an einen Bit-Computer angeschlossen: Genau wie (24) sagt.

## II.

Bringen wir das Ergebnis der bisherigen Überlegungen auf einen Punkt. Wenn ein eingetanktes Gehirn sagt: 'Mein Gehirn steckt nicht im Tank', dann hat es damit (in seiner Sprache freilich) recht. Da fragt man sich: Wieviel ist damit gewonnen? Verkennt das Gehirn im Tank nicht doch die Lage, in der es trotz allem steckt: eingetankt, also abgeschnitten von weiten Teilen der Wirklichkeit? Was hilft es, wenn das Gehirn seine wahre Lage nicht in Worte fassen kann? *Wir* sehen doch, wie es um das eingetankte Gehirn bestellt ist! (Von aussen, freilich). Können wir (von aussen) dem Gehirn wünschen, dass es sich mithilfe von Putnams Argument seiner philosophischen Sorgen entledigt? Verdient das eingetankte Gehirn denn nicht unser Mitleid? Müssen wir es nicht alarmieren, wachrütteln?

– Aber wieso? Weil es etwa im ewigen Irrtum lebt? – Worin irrt es sich denn? Vergessen wir nicht: Wir haben das eingetankte Gehirn bislang keiner einzigen falschen Aussenwelt-Meinung überführen können. Seine Beobachtungssätze sind i.A. wahr. Und das Gehirn kann sogar *wissen*, dass es mit seinen Beobachtungssätzen i.A. recht hat. Dafür sorgt Putnams Argument. Wenn also die cartesische Furcht vor dem globalen Irrtum die einzige philosophische Sorge des eingetankten

Gehirns gewesen sein sollte, dann können wir es dazu beglückwünschen, dass es diese Sorge durch Putnams Argument losgeworden ist.

Aber vielleicht sollte das eingetankte Gehirn andere philosophische Sorgen hegen als ausgerechnet die Furcht, dass all seine Aussenwelt-Meinungen falsch sind. Welche anderen Sorgen sollten das sein? Von aussen lässt sich das leicht sagen. Die Lage eines eingetankten Gehirns ist erkenntnistheoretisch und metaphysisch misslich. Sie ist erkenntnistheoretisch misslich, insofern sich sein Wissen (trotz dessen Wahrheit) nur aufs Innere des Simulationscomputers bezieht und somit weite Bereiche der Wirklichkeit nicht erfasst: Es ist also (aus unserer Sicht) dramatisch unvollständig. Und die Lage des eingetankten Gehirns ist metaphysisch misslich, insofern seine Stellung im Weltganzen auf schräge Weise beschränkt ist. Es ist in einem kleinen Ausschnitt der Wirklichkeit buchstäblich gefangen, denn es kann seinem Tank nicht entinnen.<sup>18</sup> Und schlimmer noch: Sein Schicksal hängt von Faktoren ab, die seiner Kontrolle völlig entzogen sind. Wenn die Tank-Temperatur sinkt, wenn im Computer der Strom ausfällt, wenn der Nährlösung zuwenig frische Nährstoffe zugesetzt werden, dann stirbt das Gehirn im Tank einen Tod, von dem es sich nicht einmal einen Begriff machen, dessen Ursachen es nicht erfassen und kein Stück steuern kann. Kein Zweifel: Von aussen betrachtet, von unserer Perspektive aus, steckt das Gehirn in einer ungemütlichen metaphysischen Lage.<sup>19</sup>

Diese Beobachtung liegt meiner Ansicht nach dem verbreiteten Unwillen gegen Putnams Argument zugrunde. Unser philosophischer Instinkt sagt, dass sich die eingetankten Gehirne nicht durch ein Argument abpeisen lassen sollten, das mit den schlimmstmöglichen metaphysischen Befürchtungen vereinbar ist. Und natürlich sollten wir uns genausowenig durch ein solches Argument abpeisen lassen: warnt unser Instinkt. Denn nach allem Gesagten ist es immer noch möglich, dass unsere metaphysische Lage keinen Deut besser ist als die Lage der eingetankten Gehirne. Da haben wir die metaphysische Sorge, die uns umtreibt und die von Putnams Argument nicht beruhigt wird.

---

<sup>18</sup> Sind wir nicht genauso in unserem Körper gefangen wie das eingetankte Gehirn in seinem Tank? Nun, es wird nicht mehr lange dauern, bis wir auch ausserhalb unserer Körper überleben können, nicht wahr? Die ersten Tanks sind im Bau, wir haben also bald die Wahl!

<sup>19</sup> Man mag einwenden, dass es zum Szenario eingetankter Gehirne *per definitionem* dazugehört, dass die äusseren Bedingungen des Überlebens (Tank-Temperatur usw.) ordnungsgemäss garantiert sind, so dass die drei ausgemalten Todesarten ausgeschlossen sind. Schön; aber selbst wenn es sich so verhält, bleibt das metaphysisch beunruhigende Faktum der totalen Abhängigkeit von Überlebensbedingungen, die das eingetankte Gehirn nicht im geringsten kontrollieren kann. Zum Vergleich: Im modernen Wohlfahrtsstaat haben Invalide einen Anspruch auf eine Rente, die ihr Überleben garantiert. (So zumindest die regulative Idee des Wohlfahrtsstaats). Aber dieser Anspruch ändert nichts

Das Argument kann die Sorge allenfalls mundtot machen, indem es verschiedene ihrer Formulierungen zurückweist. Aber dadurch wird es ihr nicht gerecht. Die metaphysische Sorge überlebt notfalls wortlos. Mein Ziel besteht darin, ihr die Sprache zurückzugeben. Ich möchte ihr auf angemessene Weise Ausdruck verleihen: in einer Sprache, in der sie nicht durch Putnams Argument ausgeschaltet wird. Das wird allerdings nicht einfach sein. Es wird sich zeigen, dass die aufzubietenden sprachlichen Ressourcen nicht ganz billig sind. Ich werde uns sozusagen die Folgekosten unseres philosophischen Instinkts präsentieren. Wem sie zu hoch sind, der kann am Ende seinen Instinkt immer noch kaltmachen.

Wie könnten wir unserer metaphysischen Sorge angemessen Ausdruck verleihen? Ich werde die Frage indirekt angehen und überlegen, wie ein eingetanktes Gehirn seiner analogen (und nur allzu berechtigten) Sorge angemessen Ausdruck verleihen könnte. Dieser Perspektivenwechsel hat folgenden Vorteil: Wir wissen dann, wonach wir suchen. Denn wir kennen bereits eine Beschreibung der unangenehmen metaphysischen Lage des eingetankten Gehirns: freilich in unserer Sprache und aus unserer Perspektive, von aussen. Es fragt sich: Kann das eingetankte Gehirn in seiner *eigenen* Sprache Vermutungen über die metaphysische Lage anstellen, in der es ungemütlicherweise steckt? Die Frage soll nicht mit der Frage verwechselt werden, ob das eingetankte Gehirn *herausfinden* kann, in welcher metaphysischen Lage es steckt. Gefragt ist vielmehr, ob es über sprachliche Ressourcen verfügt, seine Lage richtig zu beschreiben, ganz einerlei, ob es diese Beschreibung nun verifizieren kann oder nicht.

Natürlich könnte das Gehirn den vagen Satz sagen:

(25) Metaphysisch stimmt mit mir vielleicht irgendetwas nicht.

Und damit hätte es recht; doch das genügt nicht: Wenn diese beunruhigte Vermutung etwas austragen soll, so muss sie präziser gefasst werden. Das Gehirn muss einen informativen Grund für seine metaphysische Beunruhigung aufbieten. Was heisst informativ? Informativ genug jedenfalls wäre der Satz:

(26) Ich bin vielleicht nur ein Gehirn im Tank.

---

an ihrer Abhängigkeit von dieser Rente. Und allein diese Abhängigkeit kann sie schon beunruhigen – selbst wenn sie keine Angst vor Geldnot zu haben brauchen.

Aber wir wissen bereits, dass dieser Satz nicht zur Motivation irgendeiner metaphysischen Beunruhigung dienen kann. Er ist ja aus dem Munde eines eingetankten Gehirns genauso falsch wie aus unserem Munde. Er differenziert nicht richtig zwischen der ungemütlichen Lage im Tank und unserer (hoffentlich) komfortablen Lage ausserhalb des Tanks. Man könnte sagen: Er ist kein *Indikator* der Lage dessen, der ihn äussert. Ein Indikator wäre ein Satz, der im Munde eines eingetankten Gehirns wahr ist, aber falsch ist, wenn wir ihn äussern (oder doch *hoffentlich* falsch ist, wenn wir ihn äussern; und jedenfalls falsch ist, wenn er von einem Sprecher geäussert wird, der in keiner ungemütlichen metaphysischen Lage steckt).

Einen Indikator in diesem Sinn haben wir schon kennengelernt:

(25) Metaphysisch stimmt mit mir vielleicht irgendetwas nicht.

Aber wie wir festgestellt haben, ist dieser Indikator nicht informativ genug. Was wir dem eingetankten Gehirn an die Hand geben möchten, ist ein *informativer Indikator*, also eine gelungene Mischung aus (25) und (26).

Kann es einen solchen informativen Indikator geben? Bislang ist die Frage offen. Putnams Argument hat nur gezeigt, dass informative Sätze wie (26) nicht als Indikatoren infragekommen. Lässt sich das Argument auf alle informativen Sätze übertragen? Ich glaube nicht. Bevor ich aber einen Versuch erörtere, das Argument zu verallgemeinern, möchte ich zwei Kandidaten für informative Indikatoren zurückweisen.

Der erste Kandidat möchte das mangelnde physikalische Verständnis des eingetankten Gehirns ausschlichten.<sup>20</sup> Die Grundidee leuchtet ein: Das eingetankte Gehirn mag zwar irgendwelche "physikalischen" Gesetzmässigkeiten entdecken, die alle Vorgänge im Universalspeicher richtig beschreiben. Nun handeln diese Gesetzmässigkeiten nur von der programmgesteuerten Manipulation irgendwelcher Nullen und Einsen. Aber erstens folgt die Welt fundamental anderen physikalischen Gesetzen und zweitens besteht sie in Wirklichkeit aus anderen Materie-Bausteinen, als von der Tank-Physik beschrieben werden. (Schon der Simulationscomputer besteht nicht allein aus abgespeicherten Zahl-Sequenzen).

Eine informative Idee – eignet sie sich als Indikator? Versuchen wir es mit folgender Formulierung:

---

<sup>20</sup> So versteht Putnam Wrights Widerspruch gegen sein Argument (vergl. Wright [PPTW]:240 und Putnam [CWoB]:286).

(27) Vielleicht bestehen die materiellen Dinge aus ganz anderen Grundbausteinen, als ich bislang annahm.<sup>21</sup>

Was halten wir von einem solchen Satz in unserer Sprache? Er kann als plausible Warnung vor der Fehlbarkeit unserer wissenschaftlichen Bemühungen verstanden werden.<sup>22</sup> In der Tat, es könnte sich eines Tages herausstellen, dass Elementarteilchen oder Quarks genausowenig zur umfassenden Erklärung der physikalischen Phänomene taugen wie der (inzwischen abgedankte) Äther. Einen Fallibilismus dieser Art sollten alle zeitlichen und endlichen Wesen akzeptieren.

Weil aber nach Annahme dem eingetankten Gehirn genau dieselben physikalischen Beobachtungen vorsimuliert werden, die wir in unseren Labors machen, kann sich die fallibilistische Warnung für eingetankte Gehirne nur dann als berechtigt erweisen, wenn sie sich auch für uns als berechtigt erweist. Und somit differenziert der Satz nicht zwischen unserer Lage und der Lage eingetankter Gehirne: Er ist kein Indikator fürs Eingetanktsein.

Verschärfen wir den erfolglosen Kandidaten folgendermassen:

(28) Vielleicht bestehen die materiellen Dinge aus ganz anderen Grundbausteinen, als mir selbst irgendeine empirisch ideale Physik sagen wird (also eine Physik, die alle denkbaren Beobachtungen richtig beschreibt).

Wenn dieser Satz ein Indikator für die Lage eingetankter Gehirne sein soll, dann müsste die durch ihn ausgesprochene Vermutung in der Tanksprache wahr sein. Ist sie das? So könnte man denken. Denn die Grundbausteine der Materie sind Elektronen, Protonen und Neutronen: Objekte, von denen das eingetankte Gehirn nie wird reden können, da es mit ihnen nicht auf geeignete Weise in Kontakt treten kann. – Aber natürlich ist dieser Punkt irrelevant. Er spielt für die Auswertung der Vermutung aus (28) keine Rolle: einer *tanksprachlichen* Vermutung. Wenn sie wahr sein soll, dann müssten die *bit*-materiellen Dinge *kybernetisch* aus anderen *Bit*-Grundbausteinen zusammengesetzt sein, als dem Gehirn irgendeine empirisch ideale Physik sagen wird.<sup>23</sup> Aber damit ist nicht zu rechnen. Denn die Bit-Materie *ist* kybernetisch aus Bit-Neutronen, Bit-Elektronen und Bit-Protonen

---

<sup>21</sup> Wieder nur zur Vereinfachung unterstelle ich, dass sich die Physik in erster Linie mit der Zusammensetzung der Materie aus irgendwelchen Grundbausteinen befasst. Weniger anschaulich, aber näher an der Sache wäre es, anstelle von Grundbausteinen der Materie z.B. physikalische Gesetze zu behandeln. Am eigentlichen Argument änderte sich dadurch nichts.

<sup>22</sup> Analog Putnam in [CWoB]:286.

<sup>23</sup> Analog Putnam in [CWoB]:286/7.

zusammengesetzt: aus Bit-Bausteinen, mit denen das eingetankte Gehirn in geeignetem kausalen Kontakt steht. Schon die nicht ganz ideale Gegenwartsphysik des Gehirns enthält die richtigen Annahmen über die kybernetische Zusammensetzung bitmaterieller Gegenstände. Kurz, die Vermutung aus (28) ist in der Tanksprache falsch.

Abermals hat uns die erzwungene Uminterpretation entscheidender Wörter die Möglichkeit geraubt, mithilfe dieser Wörter zwischen unserer Lage und der Lage eines Gehirns im Tank zu differenzieren. Vorhin hatten wir die Wörter 'Gehirn', 'Tank', 'Computer' und 'in' uminterpretieren müssen, als wir sie einem eingetankten Gehirn in den Mund legen wollten; nun mussten wir die Wörter 'materieller Gegenstand', 'Baustein' und 'zusammengesetzt' uminterpretieren. Beidemal zerschlug sich durch die Uminterpretation die Hoffnung, dass sich ein gegebener informativer Satz als Indikator der Lage eingetankter Gehirne eignen könnte.

Diese schöne Regelmässigkeit führt uns in Versuchung, ein allgemeines Argument gegen die Möglichkeit informativer Indikatoren auszuhecken. Es lautet so: Kandidaten für Indikatoren sind nur dann informativ, wenn sie Wörter enthalten, die eine Uminterpretation verlangen, sobald sie einem eingetankten Gehirn in den Mund gelegt werden. Im Zuge dieser Uminterpretation werden die Referenz-Objekte der fraglichen Wörter stets aus der Welt in den Universalspeicher verschoben. Aber nach Konstruktion spiegelt der Universalspeicher den Zustand der Welt bis ins kleinste Detail wider. Somit übersteht der Wahrheitswert des betrachteten Satzes den Wechsel von unserer Sprache in die Tanksprache: Der sprachliche Wechsel wird im selben Atemzug durch den Wechsel der Umgebung (Welt versus Universalspeicher) rückgängig gemacht. Wenn jedoch der Satz aus dem Mund eines Gehirns im Tank denselben Wahrheitswert trägt wie aus unserem Mund, dann kann er kein Indikator für die Lage eingetankter Gehirne sein.

Wäre das Argument überzeugend, so könnten wir uns die Suche nach informativen Indikatoren ersparen. Nun geht meiner Ansicht nach das Argument trotz seiner Schlüssigkeit von einer falschen Prämisse aus, nämlich von der Prämisse, dass *alle* entscheidenden Wörter aus informativen Sätzen beim Übergang von der Tanksprache in unsere Sprache uminterpretiert werden müssen. Zugegeben, unsere bisherigen Beispiele scheinen diese Prämisse zu stützen. Und sie scheint sogar zu erklären, warum der Indikator

(25) Metaphysisch stimmt mit mir vielleicht irgendetwas nicht,

nicht informativ ist. Das Wort 'metaphysisch' braucht offenbar beim Übergang von der Tanksprache in unsere Sprache nicht uminterpretiert zu werden. (Daher eignet sich der Satz als Indikator). Aber dass dies Wort den Wechsel unbeschadet übersteht, zeichnet es nicht positiv aus, sondern zeigt nur, wie wenig informativ das Wort ist.

Trotzdem ist die Prämisse des Arguments überzogen. Sie malt ein allzu simples Bild vom Funktionieren unserer Sprache. Nicht alle (im Sinn des Arguments) informativen Wörter referieren so, wie wir uns das bislang zurechtgelegt haben. Selbst wenn informative Wörter nicht ohne einen gewissen kausalen Kontakt referieren können, muss dieser Zusammenhang nicht in ein Korsett gezwängt werden, in dem alles gleich aussieht.

Ich behaupte: Ohne in die magische Theorie der Referenz zurückzufallen, können wir auf informative Wörter bauen, deren Semantik den Übergang von der Tanksprache in unsere Sprache unbeschadet übersteht. Ich werde solche Wörter im folgenden *semantisch stabil* nennen. Ein Wort ist demzufolge semantisch stabil, wenn es im Munde eines eingetankten Gehirns dasselbe bedeutet wie in unserem Munde (wenn es also bei der Übersetzung aus der Tanksprache unverändert wiedergegeben werden kann).

Wenn es mir gelingt, informative und zugleich semantisch stabile Wörter aufzuweisen, dann bricht damit nicht nur die Prämisse des Arguments gegen informative Indikatoren zusammen, das wir ins Auge gefasst haben. Wir dürfen darüber hinaus hoffen, mithilfe derartiger Wörter echt informative Indikatoren für die Lage eingetankter Gehirne formulieren zu können. Und tatsächlich wird uns im letzten Abschnitt dieses Papiers ein solcher informativer Indikator über den Weg laufen.

Welche tanksprachlichen Wörter sind semantisch stabil? Zuallererst natürlich die Wörter der Logik. Kein logisches Zeichen funktioniert im Munde eines eingetankten Gehirns anders als bei uns. Doch das wird uns nicht helfen: rein logisch (ohne Deskription) können wir die Lage eingetankter Gehirne nicht gehaltvoll charakterisieren. D.h. das Arsenal der Logik ist zwar semantisch stabil, aber nicht im gewünschten Sinn informativ.

Zweitens ist das zeitliche Vokabular semantisch stabil. So muss der tanksprachliche Satz:

(29) *Als sich der Tiger näherte, floh das Känguruh,*

ohne Veränderung seiner zeitlichen Konjunktion ins Deutsche übersetzt werden:



(30) *Als sich der Bit-Tiger kybernetisch näherte, floh das Bit-Känguruh kybernetisch.*

Woran liegt es, dass Wörter für zeitliche Verhältnisse im Gegensatz zu Wörtern für räumliche Verhältnisse semantisch stabil sind? Einfach: Das Verstreichen der Zeit fühlt sich im Tank genauso an wie ausserhalb des Tanks. Wenn der Computer dem eingetankten Gehirn eine siebenstündige Traumhochzeit vorsimulieren möchte, so muss die Simulation tatsächlich sieben Stunden währen. (Den Ort des Geschehens kann der Computer dagegen ohne Erlebnisverlust von der St.-Pauls-Kathedrale ins kybernetische Innere einer geeigneten Bit-Kathedrale verlegen). Zwar kann das eingetankte Gehirn die Zeit nicht mit denselben Instrumenten *messen* wie wir. Anstelle von Uhren muss es Bit-Uhren benutzen. Aber Bit-Uhren messen Stunden, die genauso lange dauern wie unsere althergebrachten Stunden.<sup>24</sup>

Das zeitliche Vokabular ist also semantisch stabil. Doch allein durch zeitliche Angaben lässt sich die Lage des eingetankten Gehirns sicherlich nicht informativ beschreiben. Sie können allenfalls Bestandteil einer informativen Beschreibung sein. (Dass sie ein nützlicher Bestandteil solcher Beschreibungen sind, werden wir noch sehen).

Die nächste Familie von Beispielen für semantische Stabilität, die ich vorschlagen möchte, wird unseren Fall deutlicher voranbringen. Ein Mitglied der Familie ist uns schon begegnet; ohne viel Aufhebens haben wir es beim Übersetzen aus der Tanksprache unangetastet gelassen. Wir haben die tanksprachliche Rede von '*anderen* Grundbausteinen der Materie' durch die Formulierung:

andere Bit-Grundbausteine der Bit-Materie

wiedergegeben. Haben wir dabei einen Flüchtigkeitsfehler begangen oder ist das Wort '*anders*' wirklich semantisch stabil? Betrachten wir das Wort in einem harmlosen Kontext:

(31) Tiger sind anders als Känguruhs.

Das kann allerlei heissen. Für uns interessant ist folgende Interpretation:

---

<sup>24</sup> Seit Einstein sind derartige Behauptungen mit Vorsicht zu geniessen. Da die kybernetischen Geschehnisse im Universalspeicher ein getreues Abbild der physikalischen Geschehnisse im Universum sind, ticken Bit-Uhren je nach kybernetischer Relativgeschwindigkeit zur Bit-Beobachterin verschieden schnell. (Das Bit-Zwillingsparadox!) Diejenigen Bit-Uhren, die sich relativ zum Bit-Körper des eingetankten Prinzenhirns überhaupt nicht kybernetisch bewegen, messen die Zeit, in der der prinzliche Bit-Körper kybernetisch altert. Diese Zeit entspricht der subjektiv erlebten Zeit des eingetankten Gehirns. Genau genommen hätte ich also oben im Text sagen sollen: "Bit-Uhren in

(32) Tiger gehören einer anderen Art an als Känguruhs.

Nehmen wir an, ein eingetanktes Gehirn äussert diesen Satz. Wie müssen wir übersetzen? Meiner Ansicht nach so:

(33) Bit-Tiger gehören einer anderen Art an als Bit-Känguruhs.

Meiner Ansicht nach ist also der Artbegriff semantisch stabil. Und seine semantische Stabilität vererbt sich auf aus ihm abgeleitete Begriffe wie:

gleichartig, verschiedenartig, andersartig, anders.<sup>25</sup>

Um der semantischen Stabilität des Artbegriffs (und seiner Verwandten) eine gewisse Anfangsplausibilität zu verleihen, möchte ich an unsere Überlegungen zur Referenz des tanksprachlichen Artnamens 'Tiger' erinnern. Wir hatten uns im Gefolge von Putnam und Kripke darauf geeinigt, dass das eingetankte Gehirn mit diesem Wort auf alle Objekte referieren kann, die dieselbe Tiefenstruktur haben wie jene Objekte, mit denen das Gehirn ursprünglich einmal in kausalem Kontakt stand, als es den Artnamen 'Tiger' prägte. Und wir hatten festgehalten, dass es hierfür nicht darauf ankommt, ob diese Identität der Tiefenstruktur vom eingetankten Gehirn festgestellt wird, sondern nur darauf, ob sie tatsächlich vorliegt. All diese Überlegungen hatten wir in *unserer* Sprache vorgebracht. Als wir nach Referenz-Objekten des tanksprachlichen Artnamens 'Tiger' gesucht haben, kam es uns mithin auf Strukturgleichheit aus unserer Sicht an. Nun gehören strukturgleiche Objekte, aus unserer Sicht, zu ein und derselben Art. Und das bedeutet, dass wir den tanksprachlichen Artnamen 'Tiger' so interpretiert haben, dass er aus unserer Sicht auf artgleiche Objekte referiert (auf Bit-Tiger). Unsere Regeln für die Interpretation tanksprachlicher Artnamen favorisieren also eine Übersetzung durch deutschsprachige Artnamen. Und wenn tanksprachliche Artnamen durch deutschsprachige Artnamen zu übersetzen sind, *dann stimmt der Artbegriff selber in beiden Sprachen überein!*

---

relativer kybernetischer Ruhe zum prinziplichen Bit-Körper messen Stunden, die genauso lange dauern wie die Stunden einer echten Uhr in relativer Ruhe zum eingetankten Gehirn."

<sup>25</sup> Eine der vielen Ambiguitäten des Worts 'anders' könnte hier Verwirrung stiften: 'anders' bedeutet (im hier verwendeten Sinn) dasselbe wie 'andersartig' oder 'von *anderer* Art'. Die hier kursiv gesetzten Vorkommnisse dieses Wortes dürfen in ihrem jeweiligen Kontext natürlich nicht abermals wie das zu analysierende 'anders' verstanden werden. (Sonst bedeutete 'anders' dasselbe wie 'von andersartiger Art'!) Vielmehr bedeutet 'anders' in *diesen* Kontexten soviel wie 'ungleich' oder 'nicht-identisch'.

Folgender Einwand drängt sich auf: Selbst wenn alle tanksprachlichen Arten wirklich auch Arten aus unserer Sicht sein sollten, so gilt die Umkehrung noch lange nicht. Zwar mögen *Bit-Tiger* sowohl aus unserer als auch aus Tanksicht ein und derselben Art angehören; weil aber das eingetankte Gehirn nicht von *echten Tigern* sprechen kann, gehören diese Dinge nur aus unserer Sicht – und nicht aus der Tanksicht – zu ein und derselben Art. Kurz, der tanksprachliche Artbegriff hat eine geringere Reichweite als unser Artbegriff und deckt sich mit diesem nur innerhalb der Grenzen des Universalspeichers.

Ich gebe zu, dass der Einwand nur zu berechtigt ist. Ich habe ihn geradezu eingeladen, denn ich habe so über den Artbegriff zu reden angefangen, als hätte ich nur *natürliche* Arten im Sinn. Zwar habe ich das Epitheton 'natürlich' in meinem Plausibilisierungsversuch sorgfältig vermieden; aber ich habe nichts gesagt, was auf irgendeinen anderen Artbegriff hingedeutet hätte. Und in der Tat: unsere natürliche Art der Tiger konstituiert aus der eingetankten Sicht keine natürliche Art. (Allenfalls eine *übernatürliche* Art).

Woran liegt das? Der Begriff der natürlichen Art ist eng mit den erklärenden Wissenschaften verbunden (den "Natur-Wissenschaften"), deren Aufgabe u.a. darin besteht, nach tiefliegenden Strukturmerkmalen zu suchen, die sich in beobachtbaren Gesetzmässigkeiten niederschlagen und den Bereich der uns prinzipiell zugänglichen Phänomene ordnen. Den Bereich der uns *prinzipiell* zugänglichen Phänomene: das sind nicht nur die Phänomene, mit denen wir (aufgrund unserer zufälligen Position in Raum und Zeit) tatsächlich in kausalen Kontakt geraten können, sondern auch die Phänomene, mit denen wir hätten in Kontakt treten können, wären wir wannanders woanders gewesen. Das bedeutet: Unser Begriff der natürlichen Art umfasst Arten von Dingen, die in unserem physikalischen Raum vorkommen; analog umfasst der tanksprachliche kleine Bruder dieses Begriffs nur Arten von Dingen, die im "physikalischen Raum" eingetankter Gehirne vorkommen: im Universalspeicher. Und also kann der Begriff der natürlichen Art nicht semantisch stabil sein.

So weit gebe ich dem Einwand recht. Aber damit ist die Geschichte nicht zuende. Denn die Naturwissenschaften haben kein Monopol auf den Artbegriff. Arten spielen in allen Bereichen menschlicher Aktivität und Kommunikation eine herausgehobene Rolle. In der Musik, im menschlichen und gesellschaftlichen Miteinander, in Mode und Mathematik, in Märchen, Mythos und Moral – überall wollen wir sortieren und bilden Arten. Es gibt alle Arten von Arten: Tonarten und Stilrichtungen, die Art des Yuppies oder des Vertrauten, des Girlie-Looks, der Primzahl, der Hexen und Nymphen, der Sünde.

Verwenden wir in jedem der genannten Gebiete ein und denselben Artbegriff? Und wenn ja, wie soll er definiert werden? Die erste Frage möchte ich bejahen; und doch möchte ich mich weigern, die zweite Frage zu beantworten. Ich halte es für keinen sprachlichen Zufall, dass in ganz disparaten Bereichen immer wieder dieselben Wörter 'Art', 'gleichartig', 'andersartig' usw. vorkommen. Unsere Sprache ist hier nicht schludrig und braucht auch nicht desambiguiert zu werden. Trotzdem können wir unseren Artbegriff nicht durch hinreichende und notwendige Bedingungen reglementieren. Es wäre beispielsweise ein Fehler anzunehmen, dass einzig und allein der naturwissenschaftliche Artbegriff unsere wahre Idee der Art richtig trifft, während die anderen Verwendungen des Artbegriffs, streng gesehen, nicht ganz wörtlich zu nehmen sind. Nein, der naturwissenschaftliche Artbegriff fixiert nur eine unter vielen Möglichkeiten, von Arten zu reden. (Genauso fixiert der Begriff des Kartenspiels nur eine unter vielen Möglichkeiten, von Spielen zu reden). Nicht aufgrund notwendiger und hinreichender Bedingungen, sondern aufgrund von Familienähnlichkeiten reden wir in allen erdenklichen Bereichen in ein und demselben Sinn von 'Arten'.

Und nun behaupte ich: Dieser Gebrauch des Wortes 'Art' ist semantisch stabil. Wenn also eingetankte Gehirne auf den verschiedensten Gebieten zu sortieren anfangen und dabei wieder und wieder von 'Arten' reden, dann meinen sie damit genau dasselbe, was wir mit dem Wort 'Art' meinen. Die Wörter 'Art', 'gleichartig', 'andersartig' usw. müssen bei der Interpretation eingetankter Gehirne homonym ins Deutsche übersetzt werden.

Wer dafür plädiert, irgendwelche Wörter aus der Tanksprache homonym ins Deutsche zu übersetzen, macht sich verdächtig, Putnams Warnung vor der magischen Theorie der Referenz in den Wind zu schlagen. Was ist von diesem Verdacht im vorliegenden Fall zu halten? Er ist übereilt. Bevor ich ihn definitiv zurückweise, möchte ich daran erinnern, warum er nicht einmal auf den ersten Blick so plausibel ist wie etwa im Fall der homonymen Übersetzung des tanksprachlichen Worts 'Tiger'. Da war die Situation völlig eindeutig: das eingetankte Gehirn ist nie mit Tigen in Berührung gekommen und kann daher ohne Magie nicht auf Tiger referieren. Im Fall der Arten steht es besser. Wie wir gesehen haben, kennt das eingetankte Gehirn einige Arten: z.B. Bit-Tiger und Bit-Elektronen. Es steht sozusagen in Kontakt mit paradigmatischen Fällen von Arten. Selbst ein homonym zu übersetzender tanksprachlicher Artbegriff kann also durch gewisse kausale Verknüpfungen startklar gemacht werden – ohne jede Zauberei.

Nun wäre es verrückt anzunehmen, dass die Referenz des Artbegriffs auf genau dieselbe Weise fixiert wird wie etwa die Referenz des Tigerbegriffs. Es wäre sowieso schon verrückt, weil das

nicht einmal für den eingeschränkten Begriff der *natürlichen* Art vernünftig wäre; und es wäre noch verrückter nach allem, was wir uns zur Familienähnlichkeit der verschiedenen Arten (auf den verschiedenen Diskursfeldern) überlegt haben.

Zum ersten Punkt. Hinsichtlich des Begriffs der natürlichen Arten kann man mir keinen Rückfall in die magische Theorie der Referenz vorwerfen. Denn ich habe zugegeben, dass dieser Begriff nicht semantisch stabil ist. (Soviel zur Beruhigung). Trotzdem wird nicht einmal seine Referenz nach demselben Schema fixiert wie etwa die Referenz des Worts 'Tiger'. Denn sehen wir nur, was geschieht, wenn wir Putnams und Kripkes Bild auf den Begriff der natürlichen Art zu übertragen probieren: Der Sprecher gerät in kausalen Kontakt mit einer paradigmatischen Auswahl von *natürlichen Arten* (anstelle von *Tigern*), in deren Anwesenheit dann das Wort 'natürliche Art' (statt: 'Tiger') geprägt wird, mit dem Ergebnis, dass dies Wort auf alle X referiert, die eine gemeinsame Struktur mit den meisten der ursprünglichen Exemplare aus der Auswahl teilen. – Dies ist (anders als im Fall der Tiger) nicht einmal ein halbwegs gelungenes Bild dessen, was bei der Festlegung der Referenz geschieht. Es ist allenfalls eine Karrikatur, in der die Kategorien durcheinandergebracht worden sind. Denn was soll es heissen, mit ein paar natürlichen Arten (im Gegensatz zu ihren Mitgliedern) in kausalem Kontakt zu stehen? Und welche Struktur soll die natürliche Art der Tiger mit der natürlichen Art der Elektronen gemeinsam haben? Etwa die Strukturgleichheit ihrer Mitglieder? Kategorienfehler winken. Nicht einmal die Referenz des Wortes 'natürliche Art' – und erst recht nicht die des Wortes 'Art' – wird also genau so fixiert wie die Referenz von Namen für einzelne natürliche Arten. (D.h. die natürlichen Arten formen keine natürliche Art).

Trotzdem funktioniert der Begriff der natürlichen Art nicht magisch. Die kausalen Voraussetzungen für sein Funktionieren sind nur komplizierter als im Fall von Namen für einzelne natürliche Arten. Die Taufe in Anwesenheit von ein paar paradigmatischen Exemplaren genügt hier nicht. Wissenschaftliche Praxis und philosophische Reflexion müssen hinzutreten. Der Sprecher muss beispielsweise damit vertraut sein, auf welche Weise Fragen der Artgleichheit angegangen werden. Und das heisst: Bevor ihm der Begriff der natürlichen Art zuerkannt werden kann, muss er in *intensiverem* kausalen Kontakt mit den Dingen gestanden haben als im Fall des Tigerbegriffs.

Soviel zum nicht-magischen Bild der Referenz des Wortes 'natürliche Art'. (Eher eine hingeworfene Skizze als ein Bild). Wenn wir uns dem umfassenderen Begriff der Art, simpliciter, zuwenden, wird die Affaire noch unübersichtlicher. Mit *manchen* Arten geraten wir geradewegs auf kausale Weise in Berührung (z.B. mit den natürlichen Arten, wie skizziert); andere Arten gewinnen wir durch Reflexion (wie in der Mathematik, wenn wir z.B. Kegelschnitte klassifizieren); wieder andere durch

ästhetische Aufmerksamkeit (z.B. bei der Klassifikation von Stilen); noch andere durch Training unserer Menschenkenntnis oder unserer moralischen Sensibilität. Und sobald wir in einen bislang unsortierten Bereich geraten, projizieren wir unseren Artbegriff dorthin. Kein Zweifel: Unser Artbegriff wird durch tausenderlei kausale Interaktionen gestaltet – und allerdings durch unsere phantastische Phantasie, immer neue Verbindungen zu entdecken. (Wir reden z.B. von barocken Personen).

Der Artbegriff funktioniert nicht durch Zauberei. Er hängt nicht nur vom Sprecher, sondern auch von der Welt ab, die ihn umgibt. Würde im Universum das nackte Chaos tanzen, so hätten wir vermutlich überhaupt keinen Artbegriff.

Nun habe ich aber behauptet, dass der Artbegriff semantisch stabil ist, dass also eingetankte Gehirne unseren Artbegriff haben. Wie ist das möglich, wo doch ihre "Welt" (der Universalspeicher) so anders organisiert ist als unsere Welt? Anders organisiert – ja; aber *genauso ordentlich* organisiert! Den eingetankten Gehirnen präsentiert sich dieselbe Vielfalt klassifikationsfähiger Phänomene wie uns. Und da sie mit demselben Eifer und demselben Ähnlichkeitsinstinkt klassifizieren wie wir, sehe ich keinen Grund für den Verdacht, dass sie hierbei einen anderen Artbegriff verwenden als wir. Was auch immer die eingetankten Gehirne zu einer Art zusammengruppieren, wird auch aus unserer Sicht artgleich sein (Bit-Tiger, Barocksonaten, Primzahlen). Schon das ist ein Indiz dafür, dass ihr Artbegriff mit unserem übereinstimmt.

Man mag abermals einwenden, dass unter unseren Artbegriff Arten fallen, mit deren Mitgliedern eingetankte Gehirne keinen Kontakt haben können. Ich antworte: Da der Artbegriff sowieso durch mannigfache Projektionen in immer neues Gelände lebt (und auch in Gelände, wo es auf kausalen Kontakt nicht ankommt), können wir die Interpretation des tanksprachlichen Artbegriffs ruhigen Gewissens in unser Universum ausdehnen. Das ist einfach nur eine weitere Projektion, die wir dem eingetankten Gehirn aus Wohlwollen zugestehen sollten.

Es ist ein Gebot der interpretativen Fairness, alle sprachlichen Intentionen eines Sprechers beim Übersetzen zu berücksichtigen. Eingetankte Gehirne verfolgen mit ihrem Wort 'Art' dieselben Zwecke wie wir: Klassifikation mithilfe von Eigenschaften, auf die es im jeweiligen Bereich ankommt. Eingetankte Gehirne würden es sich, genau wie wir, verbitten, wenn ihnen einer vorschreiben wollte, dass der Artbegriff von vornherein auf ganz bestimmte Bereiche beschränkt bleiben soll. Sie nehmen sich, so wie wir, das Recht auf einen *offenen* Artbegriff heraus: auf einen

Begriff, der so flexibel funktioniert, dass er in immer neuen Situationen stets seine treuen alten Dienste leistet.

All diese Intentionen von Tanksprachlern können wir wie folgt auf den Punkt bringen: Wenn ein Tanksprachler mit echten Tigern in Berührung geriete und sie genau genug analysierte, dann würde es sie in seiner eigenen Sprache als 'artgleich' bezeichnen wollen.<sup>26</sup>

Auch Intentionen, die wir hier in hypothetische Situationen fortgesetzt sind, müssen bei der Interpretation eines Sprechers berücksichtigt werden: sagt das Gebot der interpretativen Fairness. Und somit ist es geboten, dem eingetankten Gehirn *unseren* Artbegriff zuzuerkennen. Das tanksprachliche Wort 'Art' ist, ohne jede Magie, semantisch stabil.

Es war mühsam, diese Behauptung plausibel zu machen; doch haben wir uns für sie nicht um ihrer selbst willen abgemüht. Wir haben deshalb nach semantisch stabilen Wörtern gesucht, weil wir mit ihrer Hilfe informative Indikatoren für die Lage eingetankter Gehirne formulieren zu können hofften. Wie nah sind wir dieser Hoffnung gekommen? Recht nah, wie es auf den ersten Blick scheinen will. Zum Beispiel scheint folgender Satz als informativer Indikator infragezukommen:

(34) Vielleicht gibt es Arten, von denen ich nicht sprechen kann.

Wenn ein Gehirn im Tank diese Vermutung vorbringt, dann wird sie, möchte man meinen, von Arten wie Tigern und Elektronen wahrgemacht: das sind Arten, von denen kein eingetanktes Gehirn sprechen kann und die trotzdem unter den tanksprachlichen Artbegriff fallen. Aber entgegen diesem Anschein kann die Vermutung nicht wahr sein; sie ist nicht einmal konsistent. Denn nehmen wir an, sie wäre wahr. Dann *kann* der Sprecher mithilfe des Ausdrucks:

Arten, von denen ich nicht sprechen kann,

---

<sup>26</sup> Ich höre den Einwand, dass dieser Konditionalsatz leer sei, weil sein Vordersatz unmöglich wahr sein könne: Tanksprachler stecken definitionsgemäss *für immer* im Tank und *könnten* daher nicht mit echten Tigern in Berührung geraten. – Falsch. Ein Tanksprachler ist ein Sprecher, dessen Sprache durch hinreichend langes Tankdasein in weiten Bereichen anders funktioniert als unsere Sprache. Wer bis gestern eingetankt war und heute aus dem Tank befreit wird, spricht immer noch die Tanksprache. (Die Referenz seiner Wörter ändert sich nicht sofort). – Ist denn die Befreiung aus dem Tank überhaupt möglich? Aber ja! Durch einen gigantischen Zufall könnten die Bit-Leiber Fleisch werden.

Der Konditionalsatz oben im Text hat also guten Sinn. Er ist genauso sinnvoll wie folgender wahre Satz: "Wenn ein Tanksprachler mit echten Tigern in Berührung geriete und sie genau genug analysierte, dann würde es sie in seiner eigenen Sprache *nicht* als 'Tiger' bezeichnen wollen".



über die fraglichen Arten reden – ein Selbstwiderspruch.

Dieser Unfall lässt sich vermeiden:

(35) Vielleicht gibt es Arten, für die ich keinen Artnamen prägen kann.

Diese Vermutung ist (anders als ihre Vorgängerin) wirklich wahr, wenn sie von einem eingetankten Gehirn vorgebracht wird: Tiger und Elektronen liefern zwei Beispiele für Arten, für die das eingetankte Gehirn keinen Artnamen prägen kann, weil es nie in geeigneten Kontakt mit Exemplaren dieser Arten treten wird.<sup>27</sup>

Doch dürfte die Vermutung deshalb kein Indikator für die Lage eingetankter Gehirne sein, weil sie auch dann wahr zu sein droht, wenn sie von Sprechern vorgebracht wird, mit denen metaphysisch alles in Ordnung ist (also z.B. – wie wir hoffen wollen –, wenn sie von *uns* vorgebracht wird). Denn angesichts der bunten Artenvielfalt, für die ich plädiert habe, dürfte selbst metaphysisch bevorzugten Sprechern die Benennung gewisser Arten verschlossen sein. Vermutlich können sie schon gewisse mathematische Arten nicht benennen: aus prinzipiellem Mangel an Rechenkapazität. Und einerlei, ob es sich so verhält oder nicht: das Schicksal eines informativen Indikators sollte von solchen Feinheiten nicht abhängen.

Der in (35) verwendete Artbegriff ist also in jedem Fall zu weit. Es geht nicht um *irgendwelche* Arten, sondern um Arten von handfesten Gegenständen. Tiger und Elektronen sind handfest genug; Zahlen und Kegelschnitte nicht. – Aber was ist ein handfester Gegenstand – wenn nicht *materiell*? Und damit scheinen wir in der Patsche zu stecken. Denn die Rede von materiellen Gegenständen ist nicht semantisch stabil! Wenn ein eingetanktes Gehirn sagt:

(36) Vielleicht gibt es Arten von materiellen Gegenständen, für die ich keinen Artnamen prägen kann,

dann meint es damit in unserer Sprache folgendes:

---

<sup>27</sup> Zwar könnte der Ausdruck 'Arten, für die ich keinen Artnamen prägen kann' mit etwas gutem Willen noch zur Einführung eines Artnamens verwendet werden: Eines Artnamens für Arten. Aber es ist gewiss kein Artnamen für Tiger!

(37) Vielleicht gibt es Arten von bit-materiellen Gegenständen, für die ich keinen Artnamen prägen kann.

Und die Elektronen und Tiger, die die Vermutung aus (35) wahrgemacht haben, können nicht für die Wahrheit der Vermutung aus (37) in Anspruch genommen werden. Denn Elektronen und Tiger sind keine bit-materiellen Gegenstände. Sie sind nicht im Universalspeicher abgelegt. Auch (36) eignet sich mithin nicht als Indikator fürs Eingetanktsein.

### III.

Zeit für eine Diagnose. Immer wieder rennen wir mit dem Kopf gegen dieselbe Wand. Sobald unsere Vorschläge für informative Indikatoren hinreichend handfest werden, enthalten sie Vokabeln, die nicht semantisch stabil sind. Die Vokabeln, durch deren Schuld wir wieder und wieder gescheitert sind, stammen allesamt aus den Naturwissenschaften, insbesondere aus der Physik. Dass wir mit solchen Vokabeln nicht weiterkommen, braucht uns nicht zu verwundern. Die gesuchten informativen Indikatoren sollen eine *metaphysisch* unangenehme Lage beschreiben. Aber wir haben uns immer nur des Arsenal der Physik bedient. Wir haben versucht, mit den sprachlichen Mitteln der Physik: *Metaphysik* zu treiben.

Wieso sollte sich die Sprache der Physik zu diesem esoterischen Zweck eignen? Es gibt keinen Grund für diese Annahme; im Gegenteil, es ist damit zu rechnen, dass die Vokabeln der Physik (und ganz allgemein: der Naturwissenschaften) metaphysisch neutral sind.<sup>28</sup> Sie dienen der Beschreibung und Erklärung der uns prinzipiell zugänglichen Phänomene; dagegen betrifft die Metaphysik Fragen, die über diesen Bereich hinausgehen. Unsere bisherigen Versuche gleichen den erfolglosen Bemühungen mancher Religionsphilosophen, mit naturwissenschaftlichen Mitteln Gott auf die Spur zu kommen; aber die Frage nach Gott ist keine wissenschaftliche Frage, die sich etwa durch Schlüsse auf die beste Erklärung beantworten lässt. Ebenso wenig können unsere metaphysischen Sorgen mit den sprachlichen Mitteln der Naturwissenschaft ausgedrückt werden. Wer die religiöse oder metaphysische Sprache allzu nah an die Sprache der Naturwissenschaften anzugleichen versucht, riskiert das Ende von Religion und Metaphysik. Die beiden Gebiete

---

<sup>28</sup> Insofern Putnams interner Realismus gegen die metaphysische Missinterpretation der Physik (oder der Naturwissenschaften oder unseres Alltagsgeredes) protestiert, stimme ich dem internen Realismus zu. Meiner Ansicht nach liefert das Argument gegen die Gehirne im Tank zugleich ein starkes Argument zugunsten dieser Version (Teilthese?) des internen Realismus: Es zeigt, dass die von der Physik (den Naturwissenschaften) beschriebene materielle Wirklichkeit nicht so unabhängig von uns ist, dass fast all unsere Meinungen über sie falsch sein könnten. Aber der interne Realismus scheint darüber hinaus ein allgemeines Metaphysik-Verbot zu enthalten. Für dieses allgemeine Metaphysik-Verbot müsste Putnam gesondert argumentieren: ein Ding der Unmöglichkeit – falls ich mit den noch folgenden Ausführungen dieses Papiers recht haben sollte.

brauchen sprachliche Autonomie, wenn sie überleben sollen. Aber Autonomie ist keine Anarchie und kein Isolationismus. Trotz ihrer sprachlichen Unabhängigkeit von der Physik (und den Naturwissenschaften) müssen in beiden Gebieten die Regeln sinnvoller Rede beachtet werden. Z.B. tut man gut daran, ihre zentralen Begriffe verständlich zu machen. Aber wie? Meine These lautet, dass die Sprachen der Religion und Metaphysik nur durch ihre Verbindungen zu anderen – alltäglicheren – Sprachverwendungen und durch ihre Verbindungen zum Leben verständlich gemacht werden können.

Am Beispiel des Artbegriffs haben wir bereits gesehen, was das ungefähr heissen soll. Ohne allzu nah an den naturwissenschaftlichen Begriff der natürlichen Art heranzurücken, aber auch nicht völlig abgetrennt von diesem reglementierten Begriff, haben wir einen Artbegriff umrissen, der überall in unserem Leben eine Rolle spielt. Er eignet sich – wie wir sehen werden – sogar für Metaphysik und Religion.

Wir brauchen mehr Begriffe dieser Sorte. Zum Beispiel brauchen wir einen metaphysisch tauglichen Gegenstandsbegriff. Dass der naturwissenschaftliche Begriff des materiellen Gegenstandes für unsere metaphysischen Zwecke nichts taugt, war der Anlass für die diagnostischen Betrachtungen, mit denen wir unsere Suche nach informativen Indikatoren vorübergehend unterbrochen haben. Es liegt nahe, zugunsten des Gegenstandsbegriffs das erfolgreiche Befreiungsmanöver zu wiederholen, das uns vorhin vom naturwissenschaftlich reglementierten Begriff der natürlichen Art zum viel flexibleren Begriff der Art, simpliciter, geführt hat. So, wie wir vorhin gesagt haben: *Es gibt alle Arten von Arten*, so können wir nun sagen: *Es gibt alle Arten von Entitäten*. Genau wie der Artbegriff funktioniert auch der Begriff der Entität in allen denkbaren Bereichen des Lebens und der Kommunikation durch Familienähnlichkeit. Er ist flexibel und offen. Er kann in immer neue Kontexte projiziert werden und ist trotzdem nicht völlig unabhängig von der Welt. Wäre alles diffus und immer im Fluss, so hätten wir keinen Begriff der Entität. Aber zum Glück wimmelt es in unserem Universum von klar umrissenen Entitäten; auch der Universalspeicher (die "Welt" eingetankter Gehirne) ist voll von ihnen. Ohne Magie (sondern durch kausalen Kontakt zu bit-materiellen Gegenständen und durch wohlverstandene Projektion) gelangen eingetankte Gehirne zu demselben Begriff von Entität, der bei uns gilt. Der Begriff der Entität ist semantisch stabil – so wie der Artbegriff. Und das bedeutet: Echte Tiger und echte Elektronen gelten auch aus der Sicht eingetankter Gehirne als "Entitäten".<sup>29</sup>

---

<sup>29</sup> Mit diesem Manöver haben wir zugleich den Bereich der Quantoren eingetankter Gehirne erweitert: Sie können über alle Entitäten quantifizieren, innerhalb und ausserhalb des Universalspeichers.

Aber unser Problem von vorhin sind wir damit immer noch nicht los: Auch Zahlen und Kegelschnitte sind "Entitäten" aus der Sicht eingetankter Gehirne. (Genau wie der Artbegriff ist der Begriff der Entität für die Formulierung informativer Indikatoren zu weit). Wir müssen die gewonnene Vielfalt wieder beschränken. Wir suchen einen Gegenstandsbegriff zwischen dem allzu engen Begriff des materiellen Gegenstandes und dem allzu weiten Begriff der Entität. Der Begriff soll semantisch stabil sein. Er soll keine Zahlen oder Kegelschnitte, wohl aber Bit-Tiger *und* Tiger umfassen – und zwar auch aus der Sicht eingetankter Gehirne.

Welche Gemeinsamkeiten haben Tiger und Bit-Tiger selbst aus der Sicht eingetankter Gehirne? Im Gegensatz zu Zahlen und Kegelschnitten existieren beide Arten von Entitäten *in der Zeit*. Bezeichnen wir solche Entitäten der Kürze halber als zeitliche Entitäten. Sie erlauben Beschreibungen mithilfe von Begriffen wie 'Veränderung', 'Dauer der Existenz', 'Zeitpunkt des Entstehens' usw. Alle diese Beschreibungen kann das eingetankte Gehirn ohne Bedeutungsverlust wortwörtlich wiederholen. Denn wie wir uns vorhin klargemacht haben, ist das zeitliche Vokabular semantisch stabil. Mithin ist auch der Begriff der zeitlichen Entität semantisch stabil.

Damit können wir unseren ersten echt informativen Indikator für die Lage eingetankter Gehirne zu fassen bekommen:

(38) Vielleicht gibt es zeitliche Entitäten, die von anderer Art sind als alle materiellen Gegenstände.

Aus unserer Sicht ist diese Vermutung falsch – sofern mit uns (wie wir hoffen wollen) metaphysisch alles in Ordnung ist.<sup>30</sup> Wenn aber ein eingetanktes Gehirn die Vermutung äussert, so meint es damit folgendes:

(39) Vielleicht gibt es zeitliche Entitäten, die von anderer Art sind als alle *bit*-materiellen Gegenstände.<sup>31</sup>

---

<sup>30</sup> Für Geist/Körper-Dualisten ist diese Behauptung vorschnell. Denn ihrer Ansicht nach sind schon mentale Entitäten andersartig als alle materiellen Gegenstände; und natürlich sind viele mentale Entitäten in jedem Fall zeitlich. Für Dualisten muss unser informativer Indikator daher wie folgt verstärkt werden: 'Es gibt zeitliche Entitäten, die von anderer Art sind als alle materiellen und alle mentalen Gegenstände'.

<sup>31</sup> Da der Begriff der mentalen Entität (i.S. des Dualisten) ganz bestimmt semantisch stabil ist, lautet der dualistische Indikator aus der vorigen Fussnote nach Übersetzung aus der Tanksprache wie folgt: 'Es gibt zeitliche Entitäten, die von anderer Art sind als alle bit-materiellen und alle mentalen Gegenstände'.

Und damit hat es recht: Elektronen und Tiger sind auch aus der Tanksicht zeitliche Entitäten; aber sie sind von anderer Art als alle denkbaren bit-materiellen Gegenstände, die im Universalspeicher abgelegt sein mögen.

Wie gewünscht, differenziert also die Vermutung zwischen der metaphysisch ungemütlichen Lage eingetankter Gehirne und der bevorzugten Lage von Wesen, mit denen metaphysisch alles in Ordnung ist. Da die Vermutung diese Differenzierung unter Berücksichtigung aller Uminterpretationen leistet, die etwa aufgrund von ungewöhnlichen metaphysischen Umständen erzwungen werden, ist sie ein echter Indikator. Und sie ist informativ. Denn sie gibt an, in welcher Hinsicht sich die ungemütliche metaphysische Lage von ihrem gemütlichen Gegenstück unterscheidet: nämlich hinsichtlich der Existenz zeitlicher, aber nicht-materieller Entitäten.<sup>32</sup>

Nun haben wir solch einen informativen Indikator nicht allein deshalb gesucht, weil wir irgendwelchen eingetankten Gehirnen helfen wollten, Vermutungen über ihre wahre Lage anzustellen. Es ging und geht uns in erster Linie um unsere eigene Lage. Hinter all unseren Bemühungen steckte die Frage, ob wir nicht in einer ähnlich ungemütlichen Lage stecken könnten wie die eingetankten Gehirne. Diese Frage können wir nun genauer fassen. Es ist die Frage, ob der gefundene informative Indikator etwa auch aus unserem Munde wahr ist: Gibt es zeitliche Entitäten, die ganz anders sind als alle materiellen Gegenstände? Wäre die Antwort auf diese Frage positiv, dann wäre es um uns so ähnlich bestellt wie um die Gehirne im Tank, in deren Sprache – wie *wir* wissen – die Frage ganz sicher positiv beantwortet werden muss.

Meiner Ansicht nach haben wir nun eine (metaphysische) Hypothese gefunden, die wir zwar formulieren, aber nicht überprüfen können. Mit naturwissenschaftlichen Methoden lässt sich die Hypothese jedenfalls nicht testen. Alle zeitlichen Entitäten, die von Naturwissenschaftlerinnen gefunden werden können, sind in *Zeit und Raum*: Es sind materielle Gegenstände.<sup>33</sup>

---

<sup>32</sup> Gibt es andere informative Indikatoren? Ja; hier ist ein Beispiel, das allerdings nicht ohne sorgfältige Überlegungen zur Philosophie des Geistes wirklich startklar gemacht werden kann: 'Mein mentales Leben spielt sich vielleicht nicht in meinem Gehirn ab'. In unserer Sprache ist diese Vermutung wahrscheinlich falsch. In der Sprache eines eingetankten Gehirns besagt sie aber soviel wie: 'Mein mentales Leben spielt sich vielleicht nicht im kybernetischen Innern meines Bit-Gehirns ab'. Und diese Vermutung beschreibt die Lage des eingetankten Gehirns zutreffend!

<sup>33</sup> Kann man die Hypothese vielleicht mit apriorischen Mitteln überprüfen – etwa mit den Mitteln der Philosophie? Möglich; immerhin liess sich die Hypothese 'Ich bin vielleicht ein Gehirn im Tank' plötzlich mit rein philosophischen Mitteln ausschalten: durch Putnams Argument. Da wir die Vermutung (38) als Antwort auf Putnams Argument formuliert haben, müssten die Philosophinnen mit völlig neuartigen apriorischen Argumenten aufwarten, um die Vermutung auszuschalten. Ausgeschlossen ist das nicht.

Die Hypothese ist also naturwissenschaftlich unverwundbar. Umgekehrt sind die Naturwissenschaften unverwundbar gegen die Wahrheit der Hypothese. Selbst wenn es zeitliche, aber nicht materielle Entitäten geben sollte, bliebe die gesamte Naturwissenschaft nicht weniger wahr. Denn sie beansprucht nur, die materielle Welt, in Zeit und Raum, zu beschreiben. Dass wir die Hypothese nicht ausschalten können, führt uns also nicht in die generelle Skepsis hinsichtlich all unserer naturwissenschaftlicher Meinungen. Auch unsere vorwissenschaftlichen Aussenwelt-Meinungen tastet die Hypothese nicht an. Sie eignet sich nicht zur Motivation einer totalen Aussenwelt-Skepsis. Dennoch ist es eine radikal skeptische Hypothese. Sie führt in den anderen Zweifel, den ich zu Beginn dieses Papiers angekündigt habe. Seine erkenntnistheoretischen Auswirkungen betreffen, wie angekündigt, nicht die Berechtigung, sondern die Reichweite unseres Aussenwelt-Wissens. Wenn es, wie die Hypothese besagt, zeitliche, aber nicht materielle Entitäten gäbe, dann würden diese Entitäten von unserem Alltagswissen und von den Naturwissenschaften nicht erfasst. Anders als gedacht, beschrieben dann die Naturwissenschaften nur einen prinzipiell unvollständigen Ausschnitt dessen, was es – in der Zeit – gibt. Und da wir nicht wissen können, ob unsere radikal skeptische Hypothese (38) zutrifft, können wir auch nicht wissen, wie gut die Naturwissenschaften prinzipiell dafür gerüstet sind, die gesamte (zeitliche) Wirklichkeit zu erfassen.

Der durch unsere Hypothese (38) ausgelöste andere (nicht-cartesische) Zweifel stürzt uns keineswegs in erkenntnistheoretische Verzweiflung hinsichtlich unserer Naturwissenschaften. Vielmehr hilft er, ihren allumfassenden Anspruch einzudämmen. Er wappnet uns gegen die Hybris der Naturwissenschaften, macht uns bescheidener. Ich behaupte: Wer sich von Zeit zu Zeit in metaphysische Hypothesen vertieft, die wie (38) in den anderen Zweifel führen, gewinnt dadurch einen neuen – besseren – Blick fürs Leben. Wie der Nicht-Zweifelnde hört er in der Nacht die Nachtigall singen, wissend: es ist eine Nachtigall, und er sieht die Milchstrasse schimmern, und er weiss um ihre Existenz, und um die Existenz der Erde, seiner Freunde, seiner Hände. Aber ohne an alledem zu zweifeln, fragt er sich: Ob das alles ist? Ob es noch mehr gibt? Anderes? *Ganz* anderes?

Die Frage braucht ihn nicht zu beunruhigen. Gewiss, sie zeigt seine intellektuellen Grenzen. Aber das ist kein Schaden. Innerhalb seiner Grenzen gibt es genug zu wissen, genug fürs tägliche Leben, genug für Alltag und Wissenschaft. Doch wenn er zudem um die Grenzen seines Wissens weiss, kann er den Alltag gelassener nehmen. Die Frage ob es jenseits des Gewohnten noch etwas *ganz* anderes gibt, relativiert den Alltag. Aber sie zerstört ihn nicht. Denn ohne den Alltag hätte die Frage nach dem Andersartigen keinen Gehalt. Die Frage ist ungefährlich, aber nicht folgenlos. Sie sorgt für *angemessenen* Abstand vom Alltag.

Natürlich kann man den richtigen Abstand zum Alltag nicht auf einen Schlag und ein für allemal gewinnen. Der richtige Abstand muss immer wieder neu eingeübt werden. Droht er zu verschwinden, so kann man ihn durch Reflexion und Meditation über radikal skeptische Hypothesen wie (38) wieder vergrößern. Solche skeptischen Hypothesen dienen dann nicht der Stützung irgendwelcher philosophischer Theorien, sondern einem praktischen Ziel. Sie bieten allerdings nicht das einzige Mittel zu diesem Ziel; andere meditative Mittel mögen denselben oder einen ähnlichen Effekt zeitigen.<sup>34</sup> Statt über unsere metaphysische Lage und die Reichweite unseres Wissens in Zweifel zu geraten und dadurch einen gewissen Abstand zum Alltag zu gewinnen, können wir uns auch auf die Endlichkeit unseres Lebens konzentrieren oder auf den verschwindend kleinen Raum, den wir im Weltall einnehmen. Die Meditation über den Tod oder über die eigene Bedeutungslosigkeit im Naturgeschehen ist eine altehrwürdige philosophische Technik. Sie wurde in den antiken Philosophieschulen mit dem Ziel der Verbesserung des Lebens praktiziert: als spirituelle Übung, wie Hadot das nennt.<sup>35</sup>

Meiner Ansicht nach gehört die radikal skeptische Meditation über Hypothesen wie (38) in diese Tradition der spirituellen Übungen. Ob sie geistesgeschichtlich mit dieser Tradition in Verbindung gebracht werden kann, vermag ich nicht zu entscheiden. Aber das spielt für meine These keine Rolle. Denn meine These ist nicht in erster Linie eine historische Behauptung. Die These besagt vielmehr, dass die skeptische Meditation in die Tradition der spirituellen Übungen *passt*.

Ein Vergleich mit einer der beiden genannten aus der Tradition überkommenen spirituellen Übungen mag das verdeutlichen. Wenn wir uns klar machen, welch winzigen Raum unsere Erde im Sonnensystem einnimmt, wieviele Millionen anderer Sonnen durch unsere Milchstrasse tanzen, wieviel andere Galaxien herumschwirren und wieviel Galaxienhaufen: dann können wir unsere Projekte, unsere enttäuschten Hoffnungen und unseren stolzen Erfolg in neuem Licht sehen. Absolut gesehen, zählen sie keinen Deut; sie sind allein aus unserer Sicht wichtig. Vielleicht sollten wir sie ein wenig gelassener nehmen?

In dieselbe Richtung, aber noch weiter, zielt die radikal skeptische Frage, ob es nicht jenseits der materiellen Welt noch andere zeitliche Entitäten gibt. Das gesamte Weltall wäre dann immer noch nicht das letzte Mass aller Dinge; unser Alltag hätte noch weniger absolutes Gewicht. Es käme noch deutlicher darauf an, welches Gewicht *wir selbst* ihm geben.

---

<sup>34</sup> Sogar nicht-meditative Mittel kommen in Betracht: Man kann mit mehr Abstand zu agieren üben, indem man bewusst mit mehr Abstand agiert. Meditation und Reflexion können solche praktischen Übungen vorbereiten helfen.

<sup>35</sup> Cf. Hadot [SE].



So weit die Parallelen zwischen den beiden spirituellen Übungen. Schon diese Parallelen zeigen deutlich genug, dass der andere Zweifel im "Strom des Lebens" mitreden kann: Er muss kein leerlaufendes Rad sein, das sich nutzlos dreht. Daher entgeht er dem Vorwurf, ein Stück schlechter, sinnloser Metaphysik zu sein. Der Sinn des anderen Zweifels ergibt sich aus der spezifischen Rolle, die er im Leben dessen spielt, der ihn aufwirft.<sup>36</sup> Nun habe ich diese Rolle bisher nur anhand der Parallelen zu einer weiteren spirituellen Übung charakterisiert. Wo liegen die Unterschiede zwischen dem anderen Zweifel und der wissenschaftlich informierten Meditation auf unsere kosmische Bedeutungslosigkeit?

Die Meditation auf die kosmische Bedeutungslosigkeit nährt nicht nur Bescheidenheit und Demut; sie entspringt ja einem gewissen wissenschaftlichen Übermut. Selbst wenn unser Lebensbereich im Vergleich zum kosmischen Gesamtgeschehen verschwindend winzig ist, brüsten wir uns doch mit der beeindruckenden Kraft zum wissenschaftlichen Blick aufs Ganze. Dieser Hybris entgeht die skeptische Meditation über den anderen Zweifel. Der andere Zweifel bringt keine Behauptungen vor, sondern *fragt*. Und da sich seine Frage nicht rational beantworten lässt, nährt er in uns auch hinsichtlich der Reichweite unserer Vernunft: Demut.

Trotzdem muss uns der andere Zweifel nicht pessimistisch stimmen. Denn seine Frage öffnet uns den Blick für bestimmte Möglichkeiten. Es könnten *gute* Möglichkeiten sein. Ich habe während unserer Suche nach informativen Indikatoren für eingetankte Gehirne immer so getan, als müssten wir *hoffen*, dass die fraglichen Indikatoren aus unserer Sicht falsch sind. Und ich habe die Lage derer, aus deren Sicht sie wahr sind, also z.B. die Lage eingetankter Gehirne, metaphysisch ungemütlich genannt. Diese Wertungen waren einseitig. Denn fragen wir uns nur, was schlimm daran wäre, wenn der gefundene informative Indikator aus unserer Sicht wahr wäre. Es gäbe dann zeitliche Entitäten von ganz anderer Art als alles Materielle. Das müsste keine Katastrophe sein. Sogar könnte es Hoffnung machen – religiöse Hoffnung.

Ich möchte diesen Punkt durch zwei Überlegungen verdeutlichen. Die erste Überlegung ist naiv: Wenn es zeitliche, nicht-materielle Entitäten gibt, so existieren sie in einer übernatürlichen Realität,

---

<sup>36</sup> Der cartesische Zweifel kann im Gegensatz dazu keine solche spezifische Rolle im Leben spielen. Seine Verneinung aller Wissensansprüche macht ihn fürs Leben untauglich. Ohne die katastrophale Verneinung alles Aussenwelt-Wissens könnten die radikalen Hypothesen der cartesischen Skeptikerin aber sehr wohl von spiritueller Bedeutung sein. Das gilt auch für folgendes Beispiel der Skepsis, das Wittgenstein zwei Wochen vor seinem Tod betrachtet hat: "Wie, wenn etwas *wirklich Unerhörtes* geschähe? wenn ich etwa sähe, wie Häuser sich nach und nach ohne offenbare Ursache in Dampf verwandelten; wenn das Vieh auf der Weide auf den Köpfen stünde, lachte und verständliche Worte redete; wenn Bäume sich nach und nach in Menschen, und Menschen in Bäume verwandelten" ([üG]:§513). Mit solchen Betrachtungen kann man sich auf den Tod vorbereiten.

ausserhalb der Natur: das ist der "Ort", wo Gott existieren könnte. (Er kann "dort" natürlich nicht mit wissenschaftlichen Mitteln gefunden werden; aber das heisst nicht, dass er nicht da ist!)

Die zweite Überlegung knüpft an die Schwierigkeiten an, die sich uns bei der Suche nach angemessenen Formulierungen für informative Indikatoren in den Weg gestellt haben. Das Problem lautete: Wie kann ein eingetanktes Gehirn über Wirklichkeitsbereiche reden, zu denen es keinen Zugang hat? Wer über Gott reden will, dem begegnen ähnliche Schwierigkeiten. Vielleicht können sie auch ähnlich angegangen werden? Der Schlüssel zum Erfolg bei der Suche nach informativen Indikatoren lag in der semantischen Stabilität von Wörtern wie 'andersartig'. Solche Wörter könnten uns auch bei der Rede über Gott aushelfen.

Und in der Tat: In diese Richtung weisen Versuche, sich einen Gottesbegriff zurechtzulegen, in denen man ohne positive Behauptungen über Gott auskommt. (Positive Behauptungen über Gott können dieser Sicht zufolge nur einem anthropomorphistischen Missverständnis entspringen). Auf jede positive Behauptung über Gott könnte man antworten:

(40) Nein, Gott ist anders.

Die Summe dieser Antworten könnte das Wort 'Gott' besser erklären als jede Definition mithilfe positiver Gottesattribute.<sup>37</sup>

Sehen wir uns diese Idee ein bisschen genauer an! Zuallererst weist man alle naturwissenschaftlichen und alltagsprachlichen Aussagen über Gott zurück. Damit gelangt man an einen Punkt, der uns bekannt vorkommen sollte:

(41) Gott ist von anderer Art als alle materiellen Gegenstände.

Hier gabelt sich der Weg. Soll Gott trotz aller Andersartigkeit ein zeitliches Wesen sein? Oder soll er ausserhalb der Zeit stehen? Beide Möglichkeiten führen in respektable religionsphilosophische Positionen. Weil sich die erste Möglichkeit besonders nahtlos an unsere bisherigen Überlegungen anschliesst, möchte ich ihr ein Stück weit folgen.<sup>38</sup> Wenn zeitliche Begriffe auf Gott angewendet werden dürfen, können wir (41) wie folgt konkretisieren:

---

<sup>37</sup> So ähnlich Benor [MRiM]:347-52.

<sup>38</sup> Dass ich die zweite Möglichkeit im folgenden ausklammere, soll nicht heissen, dass ich sie für unwichtig hielte. Sie ist nicht weniger wichtig als die erste Möglichkeit – sondern weniger leicht zu verstehen. Mir ist nicht klar, wie weit die beiden Positionen wirklich voneinander entfernt sind. Vertreterinnen eines zeitlichen Gottesbegriffs werden

(42) Gott ist ein zeitliches Wesen von ganz anderer Art als alle materiellen Gegenstände.

Und damit wird deutlich, dass der Glaube an Gott eine mögliche positive Antwort auf unsere radikal skeptische Frage nach der Existenz des Andersartigen bietet.

Wenn das so ist, dann könnten religiöser Glaube (oder zumindest eine bestimmte Version des religiösen Glaubens) und radikaler Zweifel (oder zumindest eine bestimmte Version des radikalen Zweifels) anders zusammenhängen, als die Philosophen immer angenommen haben. Bislang pflegte man religiösen Glauben und radikalen Zweifel gegeneinander auszuspielen. Der religiöse Glaube wurde beispielsweise als Gegenmittel gegen den cartesischen Skeptizismus eingesetzt.<sup>39</sup> Der andere – nicht-cartesische – Zweifel, für den ich in diesem Papier plädiere, kann dagegen mit dem religiösen Glauben Hand in Hand gehen; oder genauer gesagt: er kann ihm den Weg bahnen. Die Reise beginnt mit einer gewissen metaphysischen Beunruhigung: Könnte es nicht sein, dass wir in derselben Lage stecken wie eingetankte Gehirne? Nein, sagt Putnams Argument, ohne allerdings den Grund für die metaphysische Unruhe aus der Welt zu schaffen. Das Argument zwingt uns, die Unruhe anders zu artikulieren und nötigt uns dabei sprachliche Anstrengungen ab, mit denen wir nicht gerechnet hatten. Die artikulierbare Fassung der metaphysischen Unruhe klingt bereits wie eine religiöse Frage: Gibt es vielleicht ausserhalb der materiellen Welt noch etwas *ganz* anderes?<sup>40</sup> Gewiss kann man die Frage verneinen und für metaphysische Normalität votieren. Aber ein solches Votum kann nicht durch Argumente erzwungen werden; denn die skeptische Frage lässt sich nicht mit rationalen Mitteln beantworten. Die Frage weist uns also einerseits auf die Grenzen unserer Rationalität und unseres Wissens hin; und sie eröffnet uns andererseits Möglichkeiten jenseits dieser Grenzen: Wir *können* uns – ohne Verrat an der Rationalität, ohne Widerspruch gegen die

---

behaupten, dass Gott ewig existiert – ewig in der Zeit. Für sie hat die Frage: 'Wann begann Gottes Existenz?', guten Sinn. Sie antworten: 'Nie; es gab Ihn immer'. Vertreterinnen eines zeitlosen Gottesbegriffs müssen derartigen Äusserungen jeden Sinn absprechen. Für sie ist 'Ewigkeit' kein zeitlicher Begriff, sondern ein Begriff ausserhalb zeitlicher Vorstellungen. (Soll das heissen, dass die Frage nach dem Beginn von Gottes Dasein so sinnlos ist wie die Frage: 'Seit wann gibt es die Sieben?') Anhängerinnen eines zeitlichen Gottesbegriffs müssen übrigens nicht behaupten, dass sich Gott im Lauf der Zeit ändert. Sie können behaupten: 'Er ändert sich nie'. Auch diese Behauptung müssen ihre Kontrahentinnen für sinnlos erklären.

<sup>39</sup> Eine modernere Denkfigur derselben Art geht rückwärts vor und verwendet den robusten Glauben an die Existenz der Aussenwelt als Gegenmittel gegen religiöse Skepsis: Wenn wir trotz unüberwindbarer Argumente für den cartesischen Zweifel weiter an unserem Glauben an die Aussenwelt festhalten, dann dürfen wir – so die Denkfigur – genauso gut die unüberwindbaren Argumente für religiöse Skepsis in den Wind schlagen.

<sup>40</sup> So zumindest fragte die erste artikulierbare skeptische Hypothese, die wir gefunden haben: (38). Klingen *alle* artikulierbaren skeptischen Hypothesen wie religiöse Fragen? Das weiss ich nicht; auf unseren zweiten Fall (aus Fussnote 32) trifft diese Vermutung jedenfalls zu. Die skeptische Hypothese: 'Mein mentales Leben spielt sich vielleicht nicht in meinem Gehirn ab', wirft die religiöse Frage auf, ob mit dem physischen Tod meines Gehirns auch mein mentales Leben enden muss.

Wissenschaft – dazu entscheiden, die skeptische Frage nach der Existenz des Andersartigen zu bejahen.

Und plötzlich sehen wir, dass es keine metaphysische Katastrophe sein müsste, wenn die skeptische Hypothese zuträfe. Denn die Hypothese könnte, zum Beispiel, durch Gott wahrgemacht werden. Unsere metaphysische Lage wäre dann anders als gedacht; aber sie wäre nicht unbedingt unangenehm. Paradox formuliert, können wir die ursprüngliche metaphysische Beunruhigung dadurch besänftigen, dass wir ihr – auf eine bestimmte Weise – *recht geben*.<sup>41</sup> Natürlich wird man nicht dadurch religiös, dass man eine bestimmte skeptische Hypothese auf bestimmte Weise bejaht. Religiosität besteht nicht einfach in einer Antwort auf eine Frage; sie ist eine Lebenshaltung, in der z.B. so etwas wie Liebe zu Gott vorkommt. Doch umfasst Religiosität *auch* Antworten auf gewisse Fragen. Wer geeignete Fragen aufwirft, kann dadurch, vielleicht, einen ersten Schritt zur religiösen Lebenshaltung tun. Die Version des radikalen Zweifels, die ich den anderen Zweifel genannt habe, wirft eine solche geeignete Frage auf und erleichtert damit, vielleicht, den Anfang des Glaubens.

Der andere Zweifel ist nicht der einzig mögliche Weg zur Religiosität; und Religiosität ist nicht das einzig mögliche Ziel des anders Zweifelnden. Aber – und das war mein Punkt – religiöser Glaube und der andere Zweifel passen zusammen, so wie Frage und Antwort.

Man mag fragen: Ist nicht trotz allem die Haltung des Zweiflers anders als die des Religiösen? Unterscheiden sich die beiden nicht z.B. hinsichtlich ihrer Haltung zum Wissen? Die erste Frage ist so allgemein, dass ich sie gefahrlos bejahen kann. Die zweite Frage verdient dagegen ein klares Nein. Wer anders zweifelt, muss unsere Wissensansprüche hinsichtlich der Aussenwelt nicht infragestellen; vielmehr bezweifelt er die Reichweite unseres Wissens. Der religiöse Mensch wird das genauso sehen: Manche Gebiete sind unserem Wissen verschlossen; die materielle Welt ist es nicht. Daher heisst sie: Diesseits.

*Olaf Müller, Cambridge im April 1997*

---

<sup>41</sup> Wie verhält sich dies zu unseren Überlegungen von vorhin, denen zufolge die Wahrheit der skeptischen Hypothese deshalb metaphysisch misslich wäre, weil sie eine totale Abhängigkeit unseres Lebens von Faktoren nach sich zöge, die wir kein Stück kontrollieren können? – Man muss diese Abhängigkeit nicht unbedingt negativ werten; man kann sich ihr vertrauensvoll fügen. Ein wichtiger Bestandteil des religiösen Bewusstseins scheint mir die Bejahung dessen zu sein, was Schleiermacher die "schlechthinnige Abhängigkeit" genannt hat.

Literatur

- Benor, E.Z. [MRiM]: "Meaning and reference in Maimonides' negative theology". *Harvard theological review* 88 (1995):339-60.
- Clark, P. / Hale, B. (ed) [RP]: *Reading Putnam*. (Cambridge / Mass.: Blackwell, 1994).
- Hadot, P. [PaWo]: *Philosophy as a way of life. Spiritual exercises from Socrates to Foucault. Edited with an introduction by Arnold I. Davidson. Translated by Michael Chase*.
- Hadot, P. [SE]: "Spiritual exercises". In Hadot [PaWo]: 81-125.
- Kripke, S. A. [NN]: *Naming and necessity*. (Cambridge / Mass.: Harvard UP, 1980).
- Putnam, H. [CWoB]: "Crispin Wright on the brain-in-a-vat-argument". In Clark et al. (ed) [RP]:283-288.
- Putnam, H. [IWNH]: "Is water necessarily H<sub>2</sub>O?". In Putnam [RwHF]:54-79.
- Putnam, H. [MLR]: *Mind, language and reality. Philosophical papers, volume 2*. (Cambridge: Cambridge University Press, 1975).
- Putnam, H. [MoM]: "The meaning of 'meaning'". In Putnam [MLR]:215-271.
- Putnam, H. [RTH]: *Reason, truth and history*. (Cambridge: Cambridge UP, 1981).
- Putnam, H. [RwHF]: *Realism with a human face*. (Cambridge / Mass: Harvard University Press, 1990).
- Quine, W. V. O. [WO]: *Word and object*. (Cambridge / Mass.: MIT Press, 1960).
- Wittgenstein, L. [üG]: *Über Gewissheit*. In Wittgenstein [W]/8:113-257.
- Wittgenstein, L. [W]/8: *Werkausgabe, Band 8*. (Frankfurt: Suhrkamp, 1984).
- Wright, C. [oPPT]: "On Putnam's proof that we are not brains in a vat". In Clark et al. (eds.) [RP]:216-241.

## Nachbemerkung aus dem Jahr 2012

Dies ist ein Arbeitspapier, das ich im März und April 1997 während eines Forschungsaufenthalts in Harvard unter dem Eindruck von Diskussionen mit Hilary Putnam verfasst habe. Ich hatte Putnam davon überzeugen wollen, dass sein Argument gegen die Hypothese vom Gehirn im Tank zwar erkenntnistheoretisch voll und ganz ins Schwarze trifft, dass es aber eine verwandte skeptische Hypothese gibt, gegen die das Argument nichts ausrichtet. Putnam freute sich über die erste Hälfte dieses Vorhabens. Doch er leistete erheblichen Widerstand gegen dessen zweite Hälfte und machte mir ihre sprachphilosophischen Kosten klar. Ich lernte von ihm, dass ich in metaphysische Spekulationen geraten war. Und also biss ich in den sauren Apfel und mutierte zum Metaphysiker.

Das unmittelbare Ergebnis dieser Mutation finden Sie im vorliegenden Aufsatz. Ich versuchte die Grenzen der externalistischen Auffassung des Bezeichnens auszuloten und führte zu diesem Zweck einen naheliegenden Begriff der semantischen Stabilität ein: Ein Ausdruck ist semantisch stabil, wenn er bei Übersetzung aus der Tanksprache in unsere Sprache nicht verändert zu werden braucht. Angesichts des anhaltenden Widerstands von Putnam habe ich erheblichen Aufwand getrieben, um vorzuführen, wie ein Gehirn im Tank seine eigene Lage genau mithilfe semantisch stabiler Ressourcen treffend beschreiben kann.

Als ich Putnam den Aufsatz vorlegte, wechselte er die sprachphilosophischen Pferde und sattelte um auf eine kontextualistische Bedeutungstheorie à la Wittgenstein; seiner Ansicht nach ergaben die von mir aufgeworfenen metaphysischen Möglichkeiten keinen Sinn, weil sie sich im Strom des Lebens nicht bemerkbar machten.

Ich reagierte darauf doppelt. Einerseits präsentierte ich Putnam im kontextualistischen Rahmen einen neuen Aufsatz namens "Wundersame Verwandlungen", durch den ich aufzeigte, inwiefern skeptische Gedankenspiele zum gelingenden Leben beitragen. (Der Aufsatz wurde von Truls Wyller ins Norwegische übersetzt und erschien zwei Jahre später in *Parabel. Tidsskrift for filosofi og vitenskapsteori*, Vol. III, No. 1 (1999), pp. 87-117).

Andererseits verfasste ich unabhängig vom Kontextualismus ein ganzes Buch über semantische Stabilität und Metaphysik. Darin arbeitete ich die Grundgedanken des vorliegenden Arbeitspapiers im Detail aus. Es bildet den zweiten Teil meiner Habilitationsschrift (eingereicht April 2001) und erschien im Frühsommer 2003 (*Wirklichkeit ohne Illusionen*, Band 2, Paderborn). Während der Arbeit an dem Buch habe ich die Hauptthesen des Arbeitspapiers an unzähligen Orten vorgetragen, und zwar in München (LMU, 17. September 1997), Göttingen (25. November, 1997), Bonn (30. April 1998), Konstanz (2. Juli 1998), Heidelberg (11. Juli, 2000), Hannover (9. Januar 2001) und Bayreuth (18. Dezember 2001). Englische Versionen des Vortrags präsentierte ich im *Philosophy Club St. Andrews* (1. Mai 2002) und auf der Konferenz *The Limits of Knowledge* (Berlin, FU, 9. Juli 2002).

Weder bei diesen noch bei späteren Gelegenheiten hat irgendwer den zentralen Thesen zugestimmt, die ich seit 1997 verfechte; die einzige Ausnahme war Crispin Wright, der mir am 9.7.2002 nach meinem Vortrag sagte: "I think you've got it right". Bis vor kurzem fürchtete ich, in dieser Angelegenheit so gut wie alleine dazustehen. Aber das hat sich geändert, seit ich zu meiner Erleichterung bemerkt habe, dass David Chalmers unabhängig von meinen Bemühungen unter dem

Titel *Matrix as Metaphysics* eine metaphysische Diagnose zum Gehirn im Tank vorgelegt hat, die der meinen stark ähnelt; anstelle von semantischer Stabilität redet Chalmers von semantischer Neutralität. Sein Aufsatz kam wohl zunächst im Jahr 2003 auf der offiziellen Matrix-Homepage heraus, von wo er allerdings seit einiger Zeit verschwunden ist. Der Aufsatz erschien in der Zwischenzeit mehrmals, zuletzt in Chalmers' Buch *The character of consciousness* (Oxford, 2010, pp. 455-494, siehe insbes. pp. 481, 487/8). Das hat mich dazu veranlasst, jetzt doch das ursprüngliche Arbeitspapier aus dem Jahr 1997 in der Form zu veröffentlichen, in der ich es seinerzeit mit Putnam diskutiert habe: Abgesehen vom Layout und der neuen Zusammenfassung habe ich nicht in den Text eingegriffen.